Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Synthese und Untersuchung chemisch modifizierter epigenetisch relevanter DNA-Modifikationen

Eva Korytiaková
aus
Komárno, Slowakei

2022

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von *Herrn Prof. Dr. Thomas Carell* betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 24.08.2022

Eva Korytiaková

Dissertation eingereicht am 19.05.2022

1. Gutachter: Prof. Dr. Thomas Carell

2. Gutachterin: Prof. Dr. Lena Daumann

Mündliche Prüfung am 23.06.2022

Wege entstehen dadurch, dass wir sie gehen.

Franz Kafka

Pre moju najdrahšiu rodinu, Mami, Tato a Klaudika. Som Vám nekonečne vďačná a hrdá byť Vašou dcérou a sestrou. Aj keď ste ďaleko, v mojom srdci ste vždy pri mne.

PUBLIKATIONLISTE

Teile dieser Arbeit wurden bereits publizert oder auf Konferenzen vorgestellt.

Publikationen

- *"^{1}H NMR chemical exchange techniques reveal local and global effects of oxidized cytosine derivatives"* Romeo Dubini, **Eva Korytiaková**, Thea Schinkel, Pia Heinrichs, Thomas Carell, Petra Rovó, *ACS Phys. Chem Au* **2022** doi.org/10.1021/acsphyschemau.1c00050.

- *"Synthesis and Structure Elucidation of the Human tRNA Nucleoside Mannosyl-Queuosine"* Markus Hillmeier, Mirko Wagner, Timm Ensfelder, **Eva Korytiaková**, Peter Thumbs, Markus Müller, Thomas Carell, *Nat Commun.* **2021**, *12*, 7123.

- *"Intragenomic decarboxylation of 5-carboxycytidine"* Ewelina Kamińska,† **Eva Korytiaková**,† Andreas Reichl, Markus Müller, Thomas Carell, *Angew. Chem. Int. Ed.* **2021,** *60,* 23207-23211.

- *"Biomimetic iron complex achieves TET enzyme reactivity"* David Schmidl,† Niko Jonasson,† **Eva Korytiaková**,† Thomas Carell, Lena Daumann, *Angew. Chem. Int. Ed.* **2021**,*60*,21457–21463.

- *"Deformylation of 5-formylcytidine in different cell types"* **Eva Korytiaková**,† Ewelina Kamińska, † Markus Müller, Thomas Carell, *Angew. Chem. Int. Ed.* **2021**, *60,* 16869-16873.

- *"Analysis of an active deformylation mechanism of 5-formyl-deoxycytidine (fdC) in stem cells"* Alexander Schön,† Ewelina Kamińska,† Florian Schelter,† Eveliina Ponkkonen, **Eva Korytiaková**, Sarah Schiffers, Thomas Carell, *Angew. Chem. Int. Ed*. **2020**, *59,* 5591-5594.

- *“Active turnover of genomic methylcytosine in pluripotent cells”* Fabio Spada, Sarah Schiffers, Angie Kirchner, Yingqian Zhang, Gautier Arista, Olesea Kosmatchev, **Eva Korytiaková**, René Rahimoff, Charlotte Ebert, Thomas Carell, *Nat. Chem. Biol.* **2020**, *16*, 1411–1419.

† Diese Autoren haben zu gleichen Anteilen zur Publikation beigetragen.

Konferenzbeiträge

- **Vortrag** "*Investigation of active Demethylation of 5-methyl-2'-deoxycytidine.*" 5th German-French DNA repair Meeting, online, 2020.

- **Vortrag** "*Using highly sensitive MS/MS techniques for investigation of DNA changes on nucleoside level.*" SFB 1361 Meeting, online, 2020

- **Posterbeitrag** "*Investigation of active Demethylation of 5-methyl-2'-deoxycytidine.*" PSL Chemical Biology Symposium, Paris, Frankreich, 2019

- **Posterbeitrag** "*Synthesis of modified Nucleosides for the Investigation of active Demethylation of 5-methyl-2'-deoxycytidine.*" European Symposium of Organic Chemistry, Wien, Österreich, 2019.

- **Vortrag**: "*Exploring the Second Dimension of the Genetic Code.*", Boehringer Ingelheim PhD Workshop, Bieberach, Deutschland, 2019.

- **Vortrag**: "*Investigation of active Demethylation of 5-methyl-2'-deoxycytidine*.", FCI-Stipendiaten Treffen, München, Deutschland, 2018.

- **Posterbeitrag** "*Investigation of active Demethylation of 5-methyl-2'-deoxycytidine.*" European Symposium in Bioorganic Chemistry 2018, Gregynog, Wales, 2018.

- **Posterbeitrag** "*Investigation of active Demethylation of 5-methyl-2'-deoxycytidine.*" SFB 794 Meeting Kloster Irsee, Deutschland, 2017.

Weitere Publikationen

- *"Copper(I)-catalysed transfer hydrogenations with ammonia borane"* **Eva Korytiaková**, Niklas O. Thiel, Felix Pape, Johannes F. Teichert, *Chem. Commun*., **2017** *53*, 732-735.

DANKSAGUNGEN

An erster Stelle möchte ich meinem Doktorvater *Professor Dr. Thomas Carell* herzlich danken. Danke dir für die vielen spannenden und herausfordernden Projekte, an denen ich während meiner Doktorarbeit beteiligt sein durfte. Danke für dein Vertrauen in meine Arbeit und die Freiheit, die ich beim Bearbeiten der Projekte genießen durfte. Es war eine spannende und besondere Zeit, aus der ich sehr viel gelernt habe.

Frau *Prof. Dr. Lena Daumann* danke ich herzlich für die Übernahme der Zweitkorrektur dieser Arbeit.

Herrn *Markus Müller* möchte ich ganz herzlich für die stetige Unterstützung und für die zahlreichen wissenschaftlichen wie auch nicht-wissenschaftlichen Diskussionen danken. Unsere Gespräche haben mir immer weitergeholfen, wenn ich mich verloren gefühlt habe. Danke für deine Hilfsbereitschaft und ein offenes Ohr, das in vielen Lebenslagen sehr hilfreich war! Mindestens genauso möchte ich mich für die zahlreichen schönen Musikabende und die großartige Musikinspiration bedanken, die ich sehr genoßen habe.

Bei Frau *Slava Gärtner* bedanke ich mich für die Hilfe bei allen organisatorischen Aufgaben.

Kerstin Kurz danke ich für die grossartige Unterstützung bei allen meinen Fragen und die tolle Zusammenarbeit bei den DNA Verdaus. Du warst mir eine große Hilfe bei Bewältigung zahlreicher Probleme. Danke für deine liebe Art und Fürsorge und die vielen Pflanzen!

Bei dem Deformylierungsteam *Katharina Iwan, René Rahimoff* und *Angie Kirchner* bedanke ich mich für die schnelle und nette Aufnahme in das Team und Einarbeitung in die vielfältigen Fragestellungen. *René*, danke dass du mir in meinen ersten Monaten die Chemie der Nukleoside gezeigt hast und ein offenes Ohr für meine zahlreichen Fragen und Launen hattest. Ich habe so viel von dir lernen können und deine Hilfsbereitschaft zu schätzen gelernt. *Kathi*, es gibt so viel, wofür ich mich bei dir bedanken möchte. Es war mir eine Freude von dir lernen zu dürfen und an einem Projekt mit dir arbeiten zu können. Danke für die schnelle und intensive Einführung in die Triple Quad Welt, sowie für das Beantworten meiner unendlichen Fragen und deine Geduld. Genauso danke ich dir, dass du für mich in allen Lebenslagen da warst und immer einen hilfreichen Ratschlag für mich parat hattest. Du warst und bleibst immer die *Herrin jeder Lage*, dafür hast du meine unendliche Bewunderung. *Angie* dir danke ich, dass du mir einen Blick in die Welt der Zellkultur geboten hast. Deine positive Einstellung und lieben Worte waren ein wertvoller Begleiter bei Bewältigung einiger Probleme.

Bei meiner Projektpartnerin *Ewelina Kamińska* bedanke ich mich für die gemeinsame turbulente Zeit, die wir zusammen erlebt haben und gut bewältigt haben. Ich habe viel aus dieser Erfahrung gelernt, auf wissenschaftlicher, sowie sozialer Ebene, danke dir dafür.

Allen Studenten, die ich während der Bachelorarbeit oder eines Forschungsprakikums betreuen durfte, danke ich für eure wissenschaftliche Neugier und euer Engagment beim Bearbeiten der Forschungsprojekte. Es war mir eine große Freude, euch einen Teil von meinem Wissen und Enthusiasmus weitergeben zu können. Namentlich bedanke ich mich bei *Lars Urban, Michele Cuciuffo, Lukas Zeisel, Andreas Reichl, Katerina Pappa, Christina Kopylova, Markus Rösch, Katharina Scholz, Jannik Lidl* und *Lukas Möller*. *David Schmidl* und *Jasmin Maglic* danke ich für die Hilfe bei den unendlichen HPLC-Aufreinigungen. *David Schmidl* danke ich besonders für viele witzige Labormomente und Coldmirror Impressionen, die mich jedes Mal zum Lachen gebracht haben. Bei *Lukas Zeisel* bedanke ich mich für die Freundschaft, die bei unserer ersten Begegnung beim Ertönen von Solomun angefangen hat und sich durch viele Gespräche über die Höhen und Tiefen des Lebens weiterentwicklet hat. Du hast mich in vielen Momenten inspiriert und ich habe von dir viel gelernt. Beim Lonely Boy werde ich mich für immer mit Wärme im Herzen an unsere gemeinsame Laborzeit erinnern. Ich kann es nicht abwarten zu sehen, was deine Zukunft Großartiges bringt.

Eveliina Ponkkonen und *Julian Schackmann* danke, dass ich eure Masterarbeiten betreuen durfte und ihr mir euer Vertrauen geschenkt habt. Ihr habt euch bei den anspruchsvollen Themenstellungen sehr gut durchgekämpft und habt tolle Arbeit geleistet. Ich habe viel aus diesen Erfahrungen gelernt. *Julian,* danke dir für deine ansteckende Lebensfreude und Positivität, du hast mich mit deiner liebevollen und unendlich witzigen Art immer wieder beeindruckt. Seit der ersten Begegnung wusste ich, dass es was besonderes wird. Es war mir eine Ehre dich in dieser Zeit begleiten zu dürfen und ich habe sehr viel von dir gelernt.

Meinen Kooperationspartnern *Niko Jonasson* und *Romeo Dubini* danke ich für die spannenden und produktiven Diskussionen, bei denen ich über meinen Tellerrand schauen konnte und von euch immer was lernen konnte.

Ein ganz besonderer Dank an die ganz besondere Person, die mich von Anfang bis zum Ende der Doktorarbeitszeit begleitet, inspiriert und zum Lachen gebracht hat. *Dilara Özdemir* thank you so much for everything we went through together. You made it a wonderful time. You were there for me in the best and the worst moments. Knowing i have an awesome friend looking after me was an indescribable help. I am so happy we discovered our love for bouldering together and shared so many magical music evenings, which i will remember forever. Thank you for all the DiVa nights, i am very much looking forward to our time together in the future.

Allen meinen Laborkollegen *Nobuhiro Tago*, *Sarah Schiffers, Katharina Iwan, Alexander Schön, Markus Hillmeier, Eveliina Ponkkonen, Karerina Pappa* und *Andreas Reichl* danke ich für die vielen angenehmen Labortage. Danke dafür, dass ihr mir jahrelang die Musik unseres Laboralltags anvertraut habt und die besonderen Freitage mitgemacht habt.

In der gesamten Doktorarbeitszeit hatte ich das große Vergnügen, viele großartige Leute kennenzulernen und viele schöne Momente zu erleben. Bei *Michael Stadelmeier* bedanke ich mich für alle Krisengespräche bei diversen MS-Geräten, dafür dass du immer mit einer kleinen Espressotasse erste Hilfe geleistet hast. *Clemens Dialer* danke ich ganz herzlich für all die Gespräche über das Leben, die wie ein Balsam für die Seele gewirkt haben. *Leander Runtsch* danke ich für die gemeinsam verbrachte Zeit beim Bouldern oder Klettern, es war mir eine große Freude, die ersten Klettererfahrungen mit dir gemacht zu haben. *Martin Rossa* danke ich für die endlose Inspiration, wie man sein Bestes aus sich rausholen kann und deine wuderbare Motivation und Geduld bei unseren unvergesslichen Kletterausflügen. *Franziska Traube* danke ich für ein offenes Ohr bei allen Krisengesprächen an den seltsamsten Orten, sowie dafür, dass du immer an mich geglaubt hast und mein Selbsvertrauen verstärkt hast. *Sarah Schiffers* danke ich für die Unterstützung und unsere kulinarischen Treffen, die ich immer genoßen habe. Allen Mitgliedern des Arbeitskreises danke ich für die wunderschöne Zeit, die ich während meiner Doktorarbeit mit euch erleben durfte. Danke euch für den offenen Austausch, für die unvergesslichen Abende und Unterstützung in sämtlichen Bereichen des Doktorarbeitlebens. Vielen Dank an *Jonas Feldmann, Stefan Wiedeman, Simon Veth, Fabian Hernichel, Giacomo Ganazzoli, Felix Xu, Milda Nainyte, Bastien Viverge*, *Timm Ensfelder, Mirko Wagner, Matthias Kurz* und *Christina Schneider.*

Bei meinen liebsten Freunden *Iklima Oral* und *Maurice Lohmann* bedanke ich mich für die stetige Unterstützung, eure wundervollen Besuche und die emotional geladenen und mit Lachen gefüllten Sonntage. *Iko,* ich kann nicht glauben, dass es schon über zehn Jahre sind, ich freue mich auf alle unsere zukünftigen Abenteuer.

Den größten Dank möchte ich bei meinen liebsten Eltern *Iveta Korytiaková* und *Ján Korytiak* ausdrücken. Dakujem vám z celého srdca za vašu nekonečnú podporu, pochopenie a dôveru a že ste vždy verili ze to zvládnem. Vždy ste mi verili a dopriali mi slobodu pri každom rozhodnutí za posledných desať rokov. Som vám nekonečne vďačná, bez vašej podpory a lásky, by to všetko nebolo možné. *Klaudika Korytiaková*, tebe zo srdca ďakujem za dar tej najúžasnejšej sestry na celom svete. Ty si bola moj najväčší fanúšik, ďakujem ti, za tvoju podporu a aj v tých najtažších chvíľach si mi dodala silu, že to zvládnem a ukázala mi, že svet je vlastne ružovší ako sa mi niekedy zdá.

Drága Nagyi, szívből köszönök mindent, amit tanítottál nekem. Te vagy az igazi inspiráció, ami miatt erre az útra léptem. Végtelen melegséged és az emberek iránti szereteted örökre formált engem, és igyekszem minden nap mindent úgy tenni, hogy büszke legyél rám.

Bei meinem liebsten *Markus Hillmeier* bedanke ich mich für die schönsten zwei Jahre meines Lebens. Was für ein großartiges Glück dich getroffen zu haben! Deine endlose Unterstützung, Liebe und Motivation haben mich bis zu diesem Punkt gebracht. Du wusstest immer ganz genau, was ich gebraucht habe. Ich bin dir unendlich dankbar und kann es nicht abwarten, mein Leben mit dir zu teilen und unsere gemeinsame Reise durchs Leben zu genießen.

Inhaltsverzeichnis

ZUSAMMENFASSUNG I

SUMMARY III

1 EINLEITUNG 1

1.1 Epigenetik 1

1.1.1 Epigenetische Kontrolle der Genexpression durch Histonmodifikationen 2

1.2 Epigenetische Kontrolle der Genexpression durch DNA Methylierung 5

1.2.1 Mechanismus der DNA Methylierung 6

1.2.2 Mechanismen der Demethylierung von DNA 7

1.2.3 Nicht-kanonische DNA-Modifikationen hmdC, fdC und cadC 9

1.3 Aktive DNA Demethylierung – fdC und cadC als Intermediate der epigenetischen Regulation 12

1.3.1 Aktive DNA Demethylierung über Basenexzisionsreparatur von oxidierten mdC-Modifikationen 14

1.3.2 Aktive DNA Demethylierung über einen direkten C-C-Bindungsbruch 17

1.4 Nicht-kanonische DNA und RNA Modifikationen 21

2 AUFGABENSTELLUNG 24

3 VERÖFFENTLICHTE ARBEITEN 27

3.1 Deformylation of 5-formylcytidine in different cell types 27

3.2 Intragenomic decarboxylation of 5-carboxycytidine 34

3.3 Analysis of an active deformylation mechanism of 5-formyl-deoxycytidine (fdC) in stem cells 41

3.4 Biomimetic iron complex achieves TET enzyme reactivity 47

3.5 Active turnover of genomic methylcytosine in pluripotent cells 56

3.6 ^{1}H NMR chemical exchange techniques reveal local and global effects of oxidized cytosine derivatives 79

3.7 Synthesis and Structure Elucidation of the Human tRNA Nucleoside Mannosyl-Queuosine 90

4 UNVERÖFFENTLICHTE ARBEITEN 101

4.1 Untersuchug der intragenomischen Deformylierung mithilfe eines isotopenmarkierten Derivats von 5-Formyldesoxycytidin 101

4.1.1 Synthese von isotopen-markierten Formylcytidin Testmolekülen 103

4.1.2 Zellbiologische Daten des verfütterten Testmoleküls 106

4.1.3 Zusammenfassung und Ausblick 119

4.1.4 Projektbeiträge 120

4.1.5 Material und Methoden 121

4.2 Untersuchung der Funktion und Beteiligung von cadC an der aktiven DNA-Demethylierung 122

4.2.1 Synthese der cadC-Prodrugs für *in vivo* Studien 123

4.2.2 Zellbiologische Daten der Fütterungsexperimente von F-cadC 125

4.2.3 Zusammenfassung und Ausblick 129

4.2.4 Projektbeiträge 131

4.2.5 Material und Methoden 131

4.3 Einsatz von Nukleosid-Analoga für eine Struktur-Aktivitäts-Studie der direkten DNA-Demethylierung 133

4.3.1 Entwicklung und Synthese der fdC Analoga 134

4.3.2 Entwicklung und Synthese von cadC Analoga 138

4.3.3 Zellbiologische Fütterungsdaten 140

4.3.4 Zusammenfassung und Ausblick 144

4.3.5 Projektbeiträge 145

4.3.6 Material und Methoden 146

5 EXPERIMENTELLER TEIL 147

5.1 Material und Methoden chemische Synthese 147

5.2 Synthesevorschriften und Analytik 151

5.2.1 Allgemeine Reakionsvorschriften 151

5.2.2 Synthese von 5-Formyl-2'-desoxy-(1',2',3',4',5'-$^{13}C_5$,N^1,N^3-$^{15}N_2$)-cytidin (**1**) 153

5.2.3 Synthese von 5-Carboxy-2'-desoxy-2'-(*R*)-fluoro-cytidin (**15**) 169

5.2.4 Synthese von 5-Meth-/Eth-/Benzyl-oxy-2'-desoxy-2'-(*R*)-fluoro-cytidin (**17**, **18**, **19**) 175

5.2.5 Synthese von 5-Acetyl-2'-desoxy-2'-(*R*)-fluoro-cytidin (**31**) 187

5.2.6 Synthese von 5-(Methylsulfinyl)-2'-desoxy-2'-(*R*)-fluoro-cytidin (**32**) 197

5.2.7 Syntese von 5-(Cyano)-2'-desoxy-2'-(*R*)-fluoro-cytidin (**33**) 205

5.2.8 Syntese von 5-(Carboxamido)-2'-desoxy-2'-(*R*)-fluoro-cytidin (**45**) 212

5.3 Material und Methoden biochemische Experimente 215

5.3.1 Zellbiologische Verfütterungsexperimente, DNA und Nukleosid Pool Isolation 215

5.3.2 Enzymatischer DNA Verdau 216

5.3.3 LC-MS/MS Analyse von DNA-Proben 216

5.3.4 Nukleosid Quantifizierung in DNA-Proben 224

6 ABKÜRZUNGSVERZEICHNIS 226

7 LITERATURVERZEICHNIS 232

8 ANHANG 239

8.1 Supporting Information: Deformylation of 5-formylcytidine in different cell types 239

8.2 Supporting Information: Intragenomic decarboxylation of 5-carboxycytidine 290

8.3 Supporting Information: Biomimetic iron complex achieves TET enzyme reactivity 371

ZUSAMMENFASSUNG

Ein multizellulärer Organismus ist aus einer Vielzahl von ausdifferenzierten Zellen zusammengesetzt, die sich stark in ihrer Morphologie und Funktion unterscheiden, dabei aber gleichzeitig eine identische genomische Sequenz besitzen. Die Entstehung von hochspezialisierten Zellen aus einem gemeinsamen Vorläufer, der Stammzelle, ist durch eine hochorchestrierte epigenetische Regulation der Genexpression möglich. Durch chemische Modifikationen auf molekularer Ebene des Genoms, die eine örtlich- und zeitlich aufgelöste Expression bestimmter Gene regulieren, wird eine hohe Variabilität gewährleistet. Chemische Modifikationen an der C-5 Position der Nukleobase, wie etwa die Methylierung des kanonischen 2'-Desoxycytidins (dC), bilden eine weitere Infomationsebene der DNA, die eine regulatorische Funktion übernimmt. Das Vorhandensein von 5-Methyl-2'-desoxycytidin (mdC) in CpG-reichen Regionen des Genoms kann zu einer Abschaltung der Expression eines Gens führen. Die Entfernung dieser Modifikation kann wiederum zur Reaktivierung und Reexpression des ausgeschalteten Gens führen. Im Gegensatz zur DNA Methylierung ist der Prozess zur Entfernung dieser Methylgruppen aus dem Genom nicht vollständig aufgeklärt. Die Entfernung einer chemisch stabilen und reaktivitätsinerten Methylgruppe, die an ein sp^2-Kohlenstoff gebunden ist, ist ein thermodynamisch aufwändiger Prozess. Die Entdeckung von weiteren DNA-Modifikationen im genetischen Material von Säugetieren, 5-Hydroxymethyl-2'-desoxcytidin (hmdC), 5-Formyl-2'-desoxcytidin (fdC) und 5-Carboxyl-2'-desoxcytidin (cadC), die durch eine enzymatische Oxidation von mdC entstehen, warf für die Aufklärung des Mechanismus der Demethylierung von mdC neue Perspektiven auf. Als mögliche Funktion der neu entdeckten Modifikationen wurde deren Beteiligung an der aktiven Demethylierung von DNA postuliert. Durch zellbiologische Verfütterungsexperimente von chemisch modifizierten fdC-Modifikationen konnte 2017 zum ersten Mal gezeigt werden, dass fdC auf genomischer Ebene *in vivo* zu dC deformyliert.

Im Rahmen dieser Arbeit wurden sowohl 2'-fluorierte als auch isotopenmarkierte Derivate von fdC hergestellt und in zellbiologischen Verfütterungsstudien eingesetzt, um die biologische Funktion der genomischen Deformylierung von fdC näher zu beleuchten. Für die Darstellung der Testmoleküle wurden neue Syntesestrategien entwickelt, die ausschließlich über die ensprechenden 2'-Desoxyuridinintermediate ablaufen, um eine höchstmögliche Reinheit der C-5-modifizierten Nukleoside zu gewährleisten. Durch den Einsatz von 2'-F-fdC konnte *in vivo* gezeigt werden, dass die Deformylierungsreaktion in den Genomen von einer Vielzahl an Zellen stattfindet und die Rate dieser Deformylierung eine inverse Korrelation zur Replikationsrate der entsprechenden Zellart aufweist. Außerdem wurde gezeigt, dass der

Großteil des Deformylierungsprozesses in genomischer DNA innerhalb von 48 h passiert, wobei ein Teil von fdC als semi-stabile Modifikation im Genom verbleibt, was möglicherweise auf eine Interaktionen der fdC Aldehydeinheit mit Histonen zurückgeführt werden kann. Der mechanistische Ablauf der Deformylierungsreaktion wurde in Verfütterungsexperimenten von 2'-F-fdC zusammen mit einem 6-Aza-fdC Derivat untersucht. In dieser Studie konnte *in vivo* bestätigt werden, dass die im Genom verlaufende Deformylierungsreaktion von fdC zu dC mechanistisch über einen Angriff eines Helfernukleophils an die C-6 Position der Nukleobase abläuft.

Die Funktion und Beteiligung von cadC an der aktiven DNA Demethylierung wurde ebenfalls detailliert erforscht. Für die Untersuchung der genomischen Decarboxylierung von cadC *in vivo* war der Einsatz eines neuen Konzepts erforderlich. Durch die ionischen Eigenschaften der Carboxylatgruppe des cadCs unter physiologischen Bedingungen war die Zellpermeabilität des freien Nukleosids für zellbiologische Verfütterungsstudien nicht ausreichend. Außerdem müssten für einen effizienten genomischen Einbau eines Testmoleküls innerhalb der Zelle drei enzymatische Phosphorylierungsschritte ablaufen, was die biologische Verfügbarkeit des verfütterten Testmoleküls weiter reduziert. Um einen hocheffizienten zellulären Transport und Einbau ins Genom zu gewährleisten, wurde für die *in vivo* Studie der Decaboxylierung ein 2'-F-cadC Triphosphat synthetisiert und zusammen mit einem Nukleotidtransporter in der Zellkultur verfüttert. Mit diesem System gelang es zum ersten Mal eine intragenomische Decarboxylierung eines cadC-Derivats *in vivo* zu zeigen. Für ein tieferes Verständnis der genomischen Reaktivität von fdC und cadC wurde eine Reihe von chemischen Analoga mit ähnlichen stereoelektronischen Eigenschaften hergestellt und in Verfütterungsexperimenten gekoppelt mit einer massenspektrometrischen Analyse des Genoms eingesetzt. Die Nukleosidmodifikationen, die mit Aldehyd- und Carbonsäure-ähnlichen Funktionalitäten versehen wurden, wurden *in vivo* auf deren intrazelluläre und intragenomische Reaktivität untersucht und mit den epigenetisch relevanten Modifikationen fdC und cadC verglichen. Neben der cadC Reaktivität im Genom wurde der strukturelle Einfluss der Modifikation auf den DNA-Doppelstrang untersucht. Durch Einsatz von cadC enthaltenden Oligonukleotiden in NMR Studien konnte die Dynamik und strukturelle Veränderungen der doppelsträngigen DNA durch das Vorhandensein der Modifikation studiert werden.

Mit Hilfe eines synthetischen Eisen(IV)-Oxo Komplexes wurde der Mechanismus der Tet-Enzyme untersucht. Studien des Eisenkomplexes in einer Nukleosid- sowie Oligonukleotidumgebung zeigten, dass der biomimetische Eisenkomplex ein ideales Modellsystem für die Untersuchung der Tet-Enzyme darstellt, da es gelang, die natürliche genomische Reaktivät der Enzyme mit dem synthetischen Komplex zu simulieren.

SUMMARY

A multicellular organism consists of a multitude of differentiated cells that differ in their morphology and function, while at the same time, they possess an identical genomic sequence. The emergence of highly specialized cells, from a common progenitor, the stem cell, is possible through a highly orchestrated epigenetic regulation of gene expression. The high variability is provided by chemical modifications at the molecular level of the genome that regulate spatially and temporally resolved expression of specific genes. Chemical modifications, such as methylation of the canonical 2'-deoxycytidine (dC) at the C-5 position of the nucleobase, form another level of infomation of the DNA that performs a regulatory function. The presence of 5-methyl-2'-deoxycytidine (mdC) in CpG-rich regions of the genome can lead to silencing of a gene's expression. Removal of the modification can in turn lead to reactivation and reexpression of the silenced gene. Unlike DNA methylation, the process of a methyl group removal has not been fully elucidated. Removal of a chemically stable and unreactive methyl group attached to a sp^2 carbon is a thermodynamically high energy process. The discovery of additional DNA modifications in mammalian genomes, 5-hydroxymethyl-2'-deoxcytidine (hmdC), 5-formyl-2'-deoxcytidine (fdC) and 5-carboxyl-2'-deoxcytidine (cadC), which result from an enzymatic oxidation of mdC, has offered new perspectives for the mechanism of mdC removal. The involvement in active demethylation of DNA has been postulated as a possible function of the newly discovered modifications. In 2017, cell biological feeding experiments of chemically modified fdC modifications demonstrated for the first time *in vivo* that fdC deformylates to dC at the genomic level.

In this work, 2'-fluorinated as well as isotopically labeled derivatives of fdC were synthesized and used in cell feeding studies to gain a deeper understanding of the genomic deformylation of fdC. New synthetic strategies were developed for the synthesis of these test molecules, which proceed exclusively via 2'-deoxyuridine intermediates to ensure the highest possible purity of the C-5-modified nucleosides. By using 2'-F-fdC, it was shown *in vivo* that the deformylation reaction occurs in genomes of a variety of cells and its rate shows an inverse correlation to the replication rate of the corresponding cell type. Furthermore, it was shown that the majority of the deformylation process in genomic DNA happens within 48 h and a portion of fdC remains in the genome as a semi-stable modification, possibly showing epigenetic relevance through interactions of the fdC aldehyde moiety with histones. The mechanistic course of the deformylation reaction was investigated in feeding experiments of 2'-F-fdC together with a 6-aza-fdC derivative. In this study, it was confirmed *in vivo* that the genomic

deformylation reaction of fdC to dC proceeds via an attack of a helper nucleophile at the C-6 position of the nucleobase.

The function and involvment of cadC in the active DNA demethylation process was also investigated thoroughly. The study of genomic decarboxylation of cadC *in vivo* required a new approach. Due to the ionic properties of the cadC carboxylate group under physiological conditions, the cell permeability of the nucleoside is not ideal for cell feeding studies. Furthermore, for an efficient genomic incorporation of a modified nucleoside, three enzymatic phosphorylation steps must occur within the cell. To ensure highly efficient cellular transport and incorporation into the genome, a 2'-F-cadC triphosphate was synthesized for the *in vivo* study of decaboxylation. The compound was fed to cell cultures together with a nucleotide transporter. Using this system, it was possible to demonstrate intragenomic decarboxylation of a cadC derivative *in vivo* for the first time. For a deeper understanding of the genomic reactivity of fdC and cadC, a series of chemical analogues possessing similar stereoelectronic properties were prepared and used in feeding studies coupled to MS analysis of the genome. The nucleoside modifications, which were provided with aldehyde and carboxylic acid-like functionalities, were examined *in vivo* in terms of their intracellular and intragenomic reactivity and compared with the epigenetically relevant modifications fdC and cadC. In addition to genomic reactivity of cadC, the structural impact of the modification on the DNA double strand was investigated. By using cadC-containing oligonucleotides in NMR studies, the dynamics and structural changes of double-stranded DNA due to the presence of the modification were explored.

A synthetic iron(IV)-oxo complex was used to study the mechanism of Tet enzymes. Studies of the iron complex in a nucleoside, as well as oligonucleotide environment showed that the biomimetic iron complex is an ideal model system for studying the Tet enzymes, since it was possible to achieve and simulate the natural genomic reactivity of these enzymes.

1 EINLEITUNG

1.1 Epigenetik

Ein multizellulärer Organismus besteht aus einer Vielzahl von ausdifferenzierten Zellen, die alle von einer pluripotenten Stammzelle abstammen. Die Entwicklung eines Embryos, vom Zeitpunkt der Eibefruchtung bis hin zum multizellulären Organismus, ist ein stark regulierter Prozess, in dem sich die Zellen eines Embryos in ungefähr 200 histologisch und funktional unterschiedliche Zelltypen differenzieren.[1-2] Alle diese Zellen stammen von gemeinsamen Vorläufern, den embryonalen Stammzellen, und besitzen die gleiche genetische Information. Sie unterscheiden sich jedoch stark in ihrer Morphologie sowie ihrer Funktion. Nach der Befruchtung einer Eizelle mit dem Spermium, beides hochspezialisierte und ausdifferenzierte Zelltypen, wird eine totipotente Zygote ausgebildet, aus der im Laufe der Entwicklung jede weitere Zelle bis hin zum adulten Organismus entwickelt wird.[3-4] Aus der Zygote entsteht nach mehreren Zellteilungsschritten über die Morula ein pluripotenter Blastozyst, aus dem eine embryonale Stammzelle entwickelt wird. Die embryonale Stammzelle kann sich anschließend in die Keimbahn (multipotente Stammzelle) oder in eines der drei Keimblätter (ektoderm, mesoderm oder endoderm) spezialisieren.[5-7] In **Abbildung 1.1** sind die unterschiedlichen Stadien der Entwicklung einer Zygote hin zu den verschiedenen spezialisierten Zellen dargestellt.

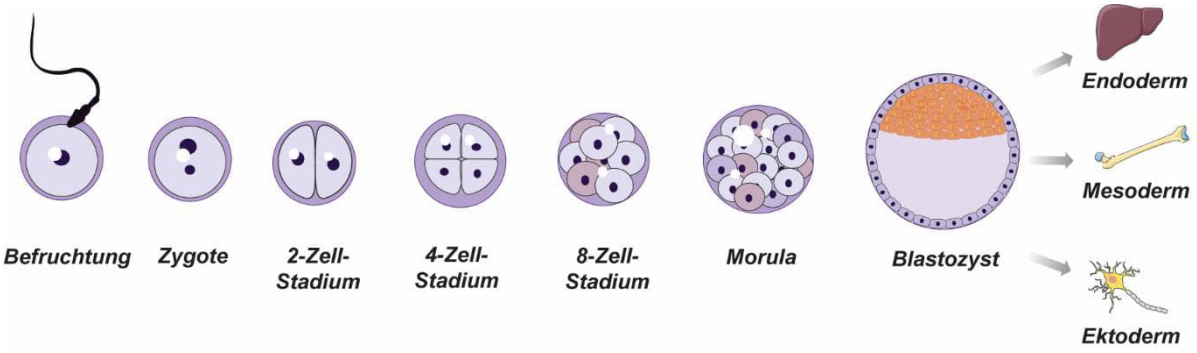


Abbildung 1.1: Übersicht der Zellentwicklung nach der Eibefruchtung mit dem Spermium zu einer totipotenten Zygote die sich über mehrere Zellteilungsschritte zu einer Morula entwickelt. Aus der Morula wird ein pluripotenter Blastozyst gebildet, welcher sich in die unterschiedlichen Keimblätter, endoderm (z. B. Leberzellen), mesoderm (z. B. Knochen) und ektoderm (z. B. Nervenzellen) entwickeln kann. Abbildung modifiziert nach Wu *et al*.[8]

Die ausdifferenzierten Zellen des Organismus besitzen alle ein identisches Genom, gleichzeitig aber ein gewebetypisches Proteom, d. h. die Zusammensetzung der Proteine unterscheidet sich stark je nach Zelltyp. Um diese große Diversität zu erreichen, muss eine multipotente Stammzelle durch exakte Regulierungsprozesse im Laufe der Entwicklung in der

Lage sein, zu definierten Zeitpunkten an distinkten Stellen des Genoms gezielt bestimmte Gene zu aktivieren, während andere Gene stillgelegt werden. Dies ist nicht nur für die Entwicklung der gewebespezifischen Proteinzusammensetzung einer Zelle, sondern auch für die Reaktion der Zellen auf äußere Umwelteinflüsse essenziell.[9] Die Fähigkeit der Regulierung der Genexpression zur Ausbildung eines gewebespezifischen Proteoms wird in der Zelle durch epigenetische Prozesse gesteuert. Dieser Differenzierungsprozess, in dem ein Genotyp während der Entwicklung einen bestimmten zellspezifischen Phänotyp ausbildet, wird als Epigenese beschrieben.[10] Diese Veränderung der Genexpression, welche nicht auf die Primärsequenz der DNA zurückzuführen ist, kann sowohl mittels Modifikation der einzelnen DNA Nukleoside als auch der mit der DNA assoziierten Proteine ablaufen. Aus chemischer Sicht handelt es sich hierbei hauptsächlich um kovalente DNA Modifikationen, sowie Modifikationen des Histonrückgrats und anderen Teilen des Chromatins.[11] Diese Modifikationen spielen eine essentielle Rolle in der Regulation der Genexpression, welche im Folgenden, beginnend bei Modifikationen des Chromatins, näher beleuchtet wird.

1.1.1 Epigenetische Kontrolle der Genexpression durch Histonmodifikationen

Chromatin bildet einen Komplex aus DNA und Proteinen, die das genetische Material ummanteln. Die Chromatinstruktur setzt sich aus einem Komplex von acht Proteinuntereinheiten, den sogenannten Histonen, zusammen. Die detailiertere Struktur dieses Histonoktamers, das auch als *Nucleosome Core Particle* (*NCP*) bezeichnet wird, konnte durch Analyse seiner Kristallstruktur aufgeklärt werden.[12-13] Das *NCP* ist aus jeweils zwei Histontetrameren zusammengesetzt, wobei ein Tetramer aus zwei Kopien der Histone H3 und H4, das andere aus je zwei Kopien von H2A und H2B besteht.[12] Jedes Histonoktamer wird in 1.65 Windungen von einem 145-147 Nukleotid (nt) langen DNA-Abschnitt spiralförmig umschlossen und bildet so das Nukleosom, die Basiseinheit des Chromatins. Die Nukleosome sind über einen kurzen (10 – 80 nt) Linker-DNA-Strang miteinander verbunden. Vier aufeinanderfolgende *NCP*s bilden die nächste Organisationsstufe des Chromatins, die sogenannte 30 nm Faser, und sind über die Linker-Histone H1 miteinander verbunden. Mittels Cryo-Elektronenmikroskopie konnte über die Struktur der 30 nm Faser gezeigt werden, dass die *NCP*s nicht nur aneinandergereiht sind, sondern sich in einer DNA-ähnlichen Doppelhelixstruktur anordnen. So organisiert sich die 30 nm-Faser letztendlich zum Chromosom.[14] Durch eine derartige strukturelle Anordnung kommt es zu einer fünf bis zehnfachen Erhöhung der Kompaktheit der DNA. Das Genom einer diploiden Zelle besteht, in Abhängigkeit vom jeweiligen Organismus, aus bis zu sechs Miliarden Basenpaaren, wobei jedes Basenpaar etwa 0.34 nm lang ist. Somit würden sich ohne eine höhere strukturelle Organisation dieses Makromoleküls in jeder Zelle nahezu zwei Meter eines

zusammenhängenden, linearen DNA-Strangs befinden. Die strukturelle Zusammensetzung des Chromatins und dessen Untereinheiten sind in **Abbildung 1.2** dargestellt.

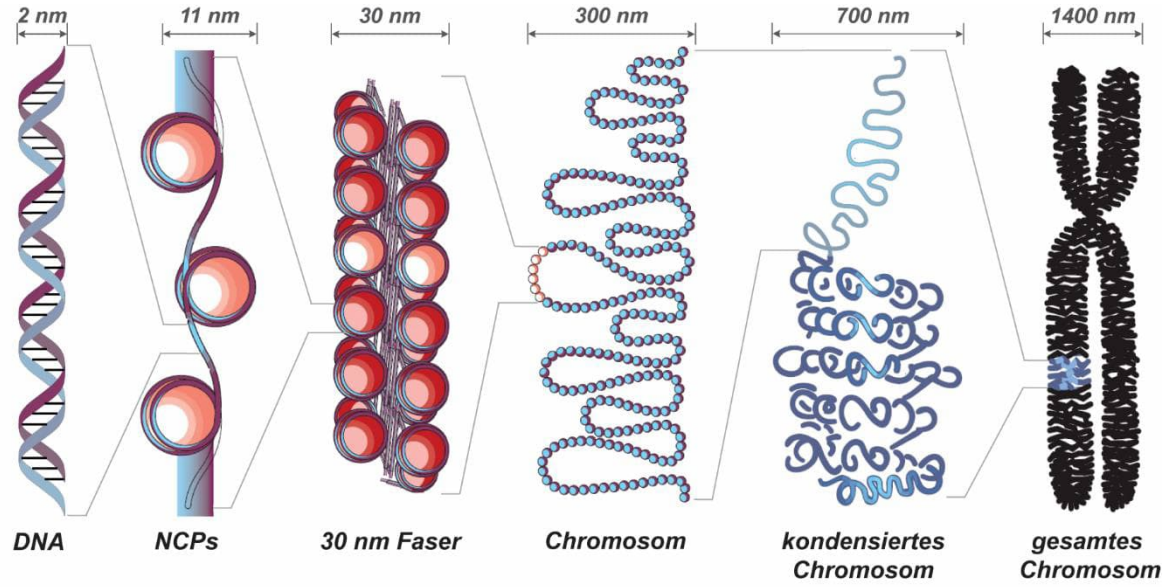


Abbildung 1.2: Strukturelle Darstellung von Chromatin sowie von dessen Untereinheiten. Das Genom wird um Histone gewickelt. Es bildet so das *Nucleosome Core Particle* (*NCP*). Vier aufeinanderfolgende *NCP*s ordnen sich in der sogenannten 30 nm Faser an, die eine Untereinheit des Chromosoms darstellt. Abbildung wurde frei erarbeitet nach *Felsenfeld et al.*[15]

Lange Zeit wurde angenommen, dass der Hauptzweck der Kompaktierung der DNA durch Histone vor allem in der Notwendigkeit der kompakten Lagerung dieses Moleküls im begrenzten zellulären Raum begründet liegt. Neuere Erkentnisse zeigen jedoch immer deutlicher, dass Histone auch eine essentielle Rolle bei der epigenetischen Regulation der Genexpression spielen.[15] Die Packungsdichte des Chromatins hat direkten Einfluss auf die Transkriptionsrate der gebundenen DNA. Die kompakte Form des Chromatins ist für Proteine nicht zugänglich und ist dadurch kein gutes Templat für biochemische Transformationen, wie Transkription, bei der die DNA Doppelhelix als Templat für die RNA Polymerase dient.[15] Die Kompaktheit und die damit einhergehende Zugänglichkeit der DNA für biochemische Reaktionen wird durch kovalente chemische Modifikationen am Chromatin beeinflusst.[16-23] Durch posttranslationale Modifikationen der *N*-terminalen Aminosäurereste der Histone kann die Zugänglichkeit der DNA für die an der Transkription beteiligten Enzyme gewährleistet werden.[21] Hierbei handelt es sich um chemische Modifizierungen in Form von Methylierung, Acylierung (Acetyl-, Propionyl- Butyryl-, Crotonyl-), Glykosylierung und Phosphorylierung (**Abbildung 1.3**). Es kommen auch komplexere Modifikationen vor, bei welchen kleinere Proteine angeknüpft werden, wie etwa Ubiquitinilierungen oder Sumoylierungen.[22-24] Die Veränderung der Chromatinpackungsdichte durch chemische Modifikationen wird besonders am Beispiel der Funktionalisierung von Lysin-Seitenketten an den Histonen deutlich. Eine

Trimethylierung der Aminogruppe einer Lysin-Seitenkette führt zur Ausbildung eines Trimethylammoniumions, und damit zu einer quarternisierung dieses Amins, welches dadurch positiv geladen vorliegt. Diese Ammonium-Gruppe kann eine anziehende elektrostatische Wechselwirkung mit dem negativ geladenen Phosphatrückgrat der DNA eingehen, was zu einer Erhöhung der Kompaktheit des DNA-Histon-Komplexes führt. Durch die erhöhte Kompaktheit wird die DNA für Proteine unzugänglicher. Die Lysine K9 und K27, die sich im Histon H3 befinden, liegen häufig trimethyliert vor. Sie stehen im Zusammenhang mit einer Inaktivierung der Genexpression. Eine Acetylierung der Aminogruppe der Lysin-Seitenkette kann in diesem Kontext einen gegenläufigen Effekt aufweisen. Durch eine Acetylierung des Amins wird das freie Elektronenpaar des Stickstoffs durch Ausbildung einer resonanzstabilisierten Amid-Struktur maskiert. Somit wird die Aminogruppe vor der Protonierung und damit vor Ausbildung einer positiven Ladung geschützt. Die Packungsdichte der Histone nimmt so ab, wodurch die DNA für Proteininteraktionen zugänglicher wird.[25]

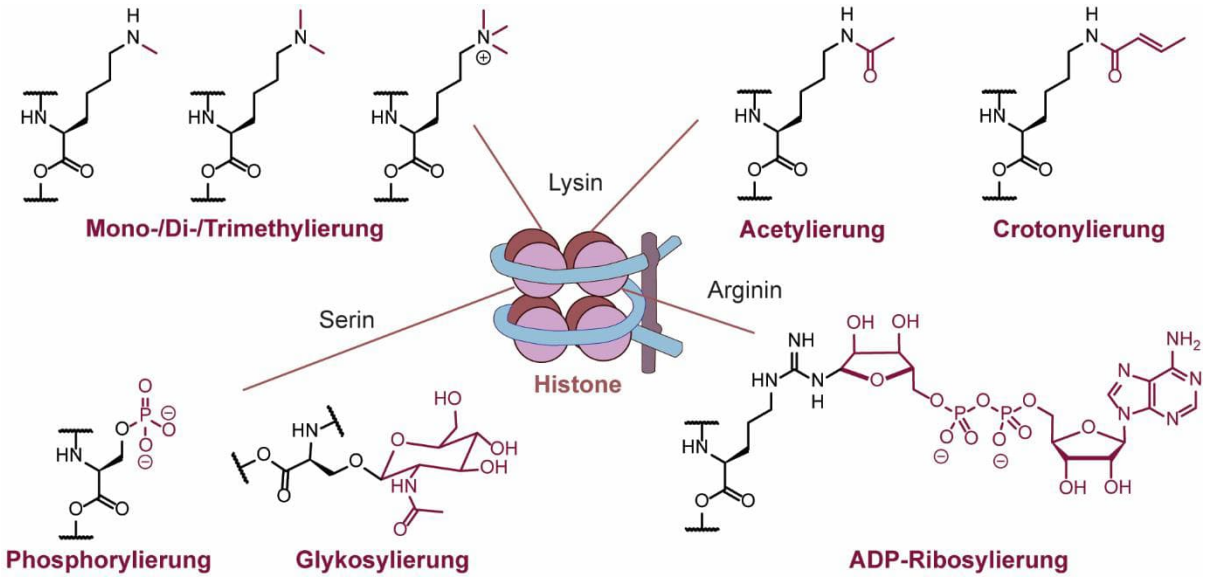


Abbildung 1.3: Ausgewählte Beispiele von postranslationalen Modifikationen der Aminosäurenseitenketten der Histonproteine von Nukleosomen. Die Aminogruppe von Lysin kann mono-, di- oder trimethyliert werden, oder mit einer Acylgruppe (Acetyl, Crotonyl) modifiziert werden. Die Hydroxygruppe von Serin kann phosphoryliert oder auch mit z. B. *O*-*N*-Acetylglucosamin glykosyliert werden. Im Fall von Arginin kann es zu einer Adenosin-Diphosphat-Ribosylierung der Seitenkette kommen.

Die Veränderung der Chromatinstruktur durch elektrostatische Kräfte, welche aus kovalenten Modifikationen der Aminosäuren resultieren, ist ein gut untersuchter Prozess. Die Abhängigkeit der Kontrolle der Genexpresion von diesem Modifikationsprozess ist jedoch sehr viel komplexer und bis heute nicht vollständig verstanden. Das Muster der posttranslationalen Histonmodifikationen, der sogenannte Histon-Code der Zelle, verändert sich dynamisch und steht im direkten Zusammenhang mit der Veränderung des Phenotyps. Die genauen Zusammenhänge sind jedoch sehr kompliziert und noch nicht eindeutig geklärt.[15] Die

posttranslational eingeführten Modifikationen an den Seitenketten der Aminosäuren in den Histonen beeinflussen sich gegenseitig und können eine Vielzahl an Reaktionen innerhalb der Zelle auslösen. Eine Monomethylierung der Seitenkette des Lysins K4 im Histon H3 führt zu Rekrutierung von Histonacetyltransferasen, welche zu einer Acetylierung von K12 im Histon H4 und dadurch zu einer Aktivierung der Transkription führen.[17, 26]

Neben der Regulierung der Genexpression in der Zelle durch die Veränderung der Chromatinstruktur können auch kovalente Modifikationen auf der Ebene des Genoms zu einer Aktivierung bzw. Deaktivierung der Genexpression führen. Diese regulatorische Funktion wird zum Großteil über die Methylierung des 2'-Desoxycytidins (dC) ausgeführt.[8, 27-32] Im folgenden Kapitel wird im Detail darauf eingegegangen, wie die Methylierung bzw. De-Methylierung der DNA zum Ein- bzw. Ausschalten von Genen führen kann.

1.2 Epigenetische Kontrolle der Genexpression durch DNA Methylierung

Eine der wichtigsten Stellschrauben für die Regulierung der Genexpression in Wirbeltieren stellt die vererbbare DNA Methylierung in Form von 5-Methyl-2'-desoxycytidin (mdC) dar.[11, 33-35] Die Methylierung der C-5 Position der Cytosinnukleobase zu mdC ist die am häufigsten vorkommende Modifikation der genomischen DNA. Innerhalb des Genoms liegen 4-6% aller Cytidine methyliert vor.[33, 35] Die Modifikation ist am häufigsten in CpG Dinukleotiden, bzw. *Clustern* von CpGs, sogenannten CpG-*islands*, zu finden. Diese werden oft mit Promoter-Regionen assoziiert und liegen zu etwa 70-80% methyliert vor.[36-37] Die Methylierung der dCs kann durch unterschiedliche Mechanismen zur Gen-Inaktivierung führen. Die methylierte Nukleobase mdC ist wie unmodifiziertes dC weiterhin in der großen Furche der DNA-Doppelhelix lokalisiert und kann dadurch mit verschiedenen Proteinkomplexen interragieren, die mdC spezifisch erkennen. Es wurden bis heute in Säugetieren sechs Proteine, die sogenannten *Methylated CpG Binding Proteins* (MBDs) identifiziert, welche mdC erkennen und zu einer spezifischen Interaktion führen: Kaiso, MECP2 und MBD1-4.[38] Für die Regulierung der Genaktivität durch die Interaktion der mdC-bindenden Proteine ist es entscheidend, wo im Genom die methylierten Cytidine lokalisiert sind. Wenn die Promotor Region eines Gens demethyliert vorliegt, wird die Expression des entsprechenden Gens stimuliert. Dieser Effekt tritt auch dann auf, wenn *downstream* des Gens CpG-Sequenzen methyliert vorliegen.[39-42] Wenn die *CpG-islands* in Promoterregionen methyliert sind, kann es durch die Bindung von MBDs zur Inaktivierung der Genexpression kommen, indem weitere Repressor-Komplexe rekrutiert werden. Für MECP2 ist zum Beispiel bekannt, dass sich ein Multikomponentenkomplex mit Histon Deacetylasen und einem aktiven Co-Repressor Protein

bildet.[43-44] Dadurch kommt es zur Veränderungen in der Chromatinstruktur was schlussendlich zur Veränderungen in der Genexpression führt.[43] Ein weiterer natürlicher, jedoch selten auftretender Geninaktivierungsweg durch DNA Methylierung wird dadurch versursacht, dass Transkriptionsfaktoren durch die DNA Methylierung an die Targetsequenz nicht binden können.[38, 45-46]

1.2.1 Mechanismus der DNA Methylierung

Die Methylierung der genomischen Cytidine erfolgt postsynthetisch durch die Familie der DNA-Methyltransferasen (Dnmts). In Säugetieren sind die Methyltransferasen Dnmt1, Dnmt3a und Dnmt3b bekannt.[38] Die charakteristischen Methylierungsmuster werden während des Entwicklungsprozesses von den *de novo* Methyltransefasen Dnmt3a und Dnmt3b unabhängig von der Replikation eingeführt.[47-49] Durch die vornehmliche Funktion von Dnmt1 als *maintenance* Methyltransferase werden Methylierungsmuster während der Replikation aufrechterhalten. Diese ist in der Lage hemimethylierte Sequenzen zu erkennen und zu methylieren, kann aber nur eingeschränkt *de novo* Modifizierung durchführen. Die strukturelle Besonderheit von Dnmt1 liegt in der Cystein-reichen, *N*-terminalen Zink-koordinierenden Strukturuntereinheit, die die *de novo* Methylierung blockiert.[50] Bei einer proteolytischen Abspaltung dieser Untereinheit kann das Enzym unmethylierte DNA erkennen und methylieren.[50] Um die genomischen Cytidine an der C-5 Position zu methylieren benötigen die Dnmts den Kofaktor *S*-Adenosylmethionin (SAM), der als Methyldonor fungiert.[51-53] In der **Abbildung 1.4** ist der postulierte mechanistische Ablauf der Methylierung von dC zu mdC auf molekularer Ebene dargestellt. Nach der Erkennung der zu methylierenden Stellen im Genom durch Dnmts erfolgt im ersten Schritt ein Michael-artiger nukleophiler Angriff einer Cysteinseitenkette des aktiven Zentrums des Enzyms an die C-6 Position der Cytosinbase. Gleichzeitig kommt es zur Protonierung von *N*-3 der Nukleobase durch eine koordinierende Glutamatseitenkette, was zu einer starken Aktivierung der C-5 Position führt.[54] Der nukleophil aktivierte C-5 Kohlenstoff greift im Anschluss nach erneuter Deprotonierung die Methylgruppe von SAM an, wobei eine C5-C6-gesättigte Cytosin Spezies und *S*-Adenosylhomocystein (SAhC) gebildet werden. Nach *syn* Eliminierung des C-5 Protons wird Cystein eliminiert und das methylierte Nukleosid mdC entsteht unter Rearomatisierung des Cytosinrings und gleichzeitiger Regeneration des Enzyms. Der mechanistische Ablauf der dC Methylierung wurde unabhängig voneinander in zwei Arbeiten von *Verdine*[51] und *Newman*[55] gezeigt. In beiden Studien wurde 5-fluoriertes dC in Oligonukleotide eingebaut und an prokaryotische[51] oder menschliche Zellen[55] verfüttert. Durch die Fluorierung der Base konnte es zu keiner Eliminierung kommen, wodurch die Dissoziation von DNA und Enzym verhindert wird. Der so

gebildete intermediäre DNA-Enzym Komplex konnte auf diesem Wege isoliert und charakterisiert werden.

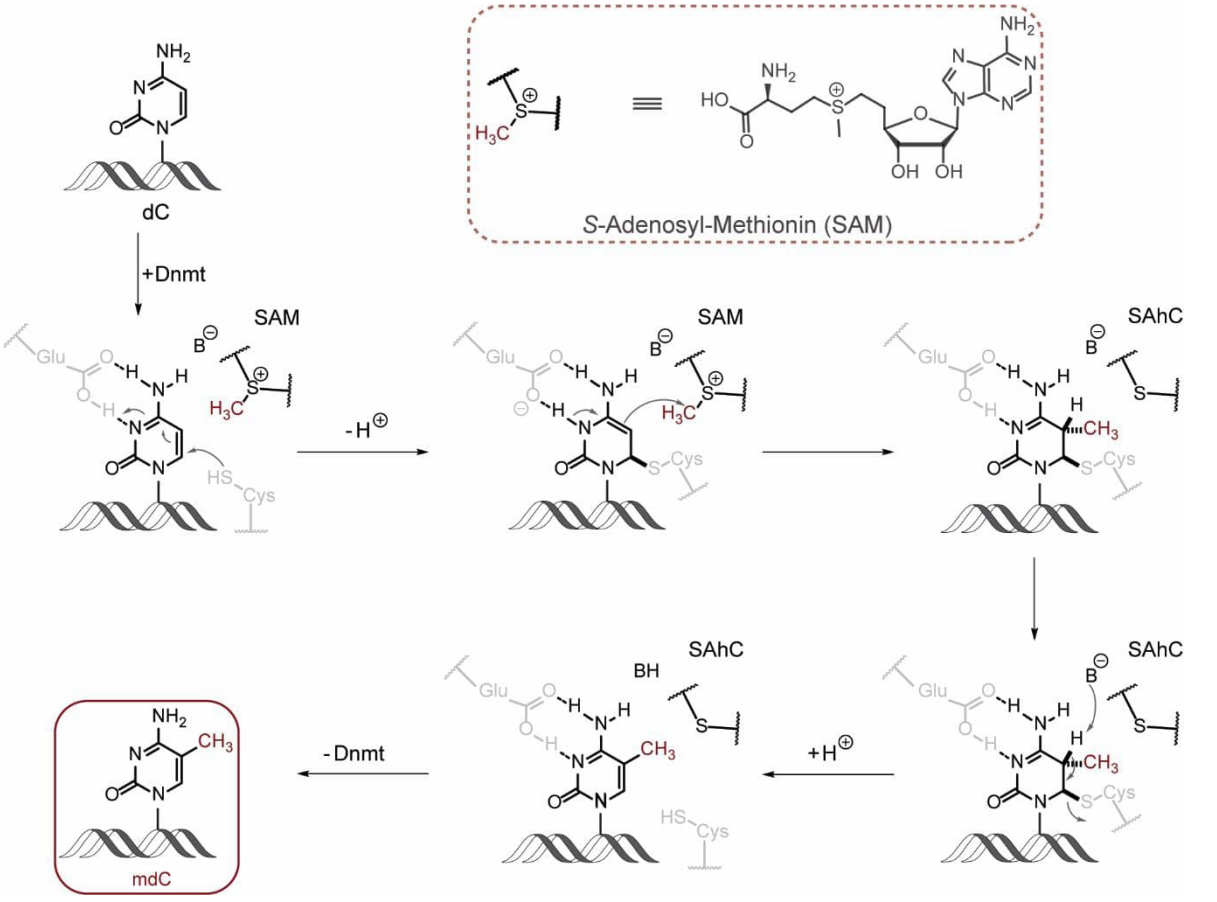


Abbildung 1.4: Postulierter mechanistischer Ablauf der Dnmt-katalysierten Methylierung von genomischem dC zu mdC mit SAM als Kofaktor. Ein Michael-artiger Angriff einer Cysteinseitenkette im aktiven Zentrum des Enzyms führt zur Aktivierung der C-5 Position der Nukleobase. Im Anschluss erfolgt ein nukleophiler Angriff des aktivierten Cytosin-C-5 Kohlenstoffs an die elektrophile Methylgruppe von SAM, was zu einer Methylierung der Nukleobase und zur Sättigung der C-5-C-6-Bindung führt. Eine anschließende Deprotonierung an C-5 und Rearomatisierung des Heterozyklus unter Freisetzung der katalytischen Spezies (Cystein) führt zur Ausbildung von mdC. (SAM = *S*-Adenosylmethionin; SAhC = *S*-Adenosylhomocystein)

1.2.2 Mechanismen der Demethylierung von DNA

Die zunächst als stabiler epigenetischer Marker gesehene Methylierung von dC in DNA zu mdC zeigte sich über Jahre intensiver Forschung als nicht so statisch wie ursprünglich gedacht.[8] Bei der stark regulierten Genaktivierung, sowie der Reaktivierung stillgelegter Gene spielt die Demethylierung von mdC in der zellulären Entwicklung eine zentrale Rolle. Der Zusammenhang zwischen Deaktivierung von Genen und Methylierung von DNA ist lange bekannt, jedoch sind längst nicht alle Mechanismen aufgeklärt, wie die Demethylierung von mdC abläuft. Eines der Modelle, welches die Entfernung von mdC aus dem Genom erklärt, ist die passive Demethylierung von mdC, bei der die mdC Konzentration in der DNA über mehrere Replikationsschritte verdünnt wird.[56] Während der Replikation paart die methylierte CpG-

Stelle des Templat-Strangs mit einer nicht-modifizierten CpG-Stelle des neu synthetisierten Strangs. Dies resultiert in einem hemi-methylieren CpG-Motiv, welches durch den *maintanence* Methyltranserasenkomplex Dnmt1/UHRF1 erkannt und gebunden wird. Der nicht methylierte Teil der hemimethylierten Stelle wird methyliert, um das symmetrische Methlyierungsmuster aufrecht zu erhalten.[57] Durch Fehlen oder Inaktivität von Dnmt1, oder aufgrund des Vorhandenseins der oxidierten mdC Derivate werden hemimethylierte Stellen im Genom nach der Replikation nicht wieder komplementär methyliert, was im Verlauf mehrerer Zellteilungsschritte zum Verlust der epigenetischen Information führt. Diese mdC Verdünnung aus dem Genom wird durch die passive Demethylierung beschrieben und ist schematisch in **Abbildung 1.5** dargestellt.[57]

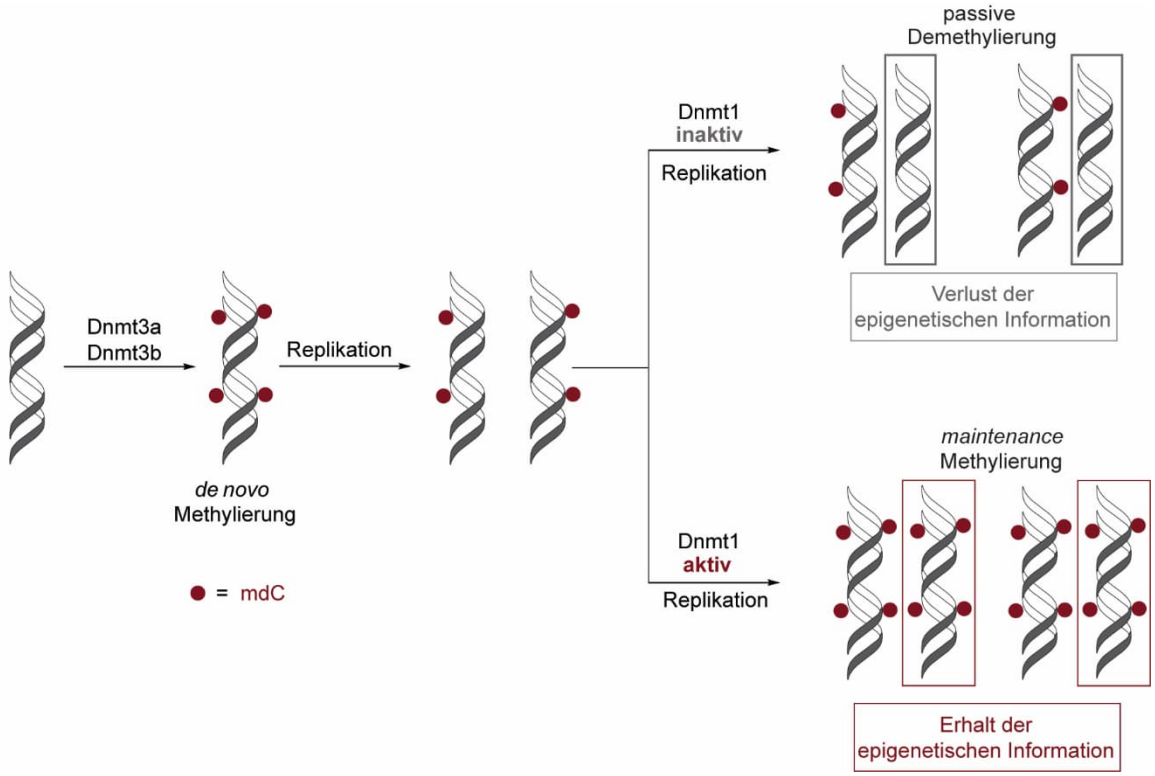


Abbildung 1.5: Eine schematische Darstellung der passiven Demethylierung, die im Verlauf mehrerer Replikationsschritte zum Verlust der epigenetischen Information führt. Ein DNA-Strang wird von den *de novo* Methyltransferasen Dnmt3a und Dnmt3b methyliert. Nach einem Replikationschritt entstehen in den neu gebildeten Strängen hemimethylierte Stellen. Bei Inaktivität von Dnmt1 kommt es in weiteren Replikationsschritten zur weiteren Verdünnung und Verlust der Methylierung (oberer Teil der Abbildung). Wenn Dnmt1 aktiv ist werden hemimethylierten Stellen erkannt, modifiziert (unterer Teil der Abbildung) und die epigenetische Information bleibt erhalten.

Die DNA Methylierung stellt zwar eine vererbbare Langzeitkontrolle der Geninaktivität dar, jedoch wird im Genom und auf genspezifischer Ebene Demethylierungsaktivität beobachtet.[8] Mit dem Modell der passiven Demethylierung kann nicht erklärt werden, wie mdC kurz nach der Befruchtung aus dem Genom in der beobachteten Rate entfernt wird. Wie aus der **Abbildung 1.6** zu entnehmen ist, kommt es während Embryogenese zu einem drastischen

Abfall der Konzentration von mdC sowohl im väterlichen als auch im mütterlichen Genom. Dieser mdC Verlust kann im paternalen Genom nicht durch eine passive, replikationsbedingte Verdünnung von mdC erklärt werden, da dieser Effekt bereits vor der ersten Replikation auftritt.[8, 58] Der beobachtete Rückgang der mdC Menge ist so drastisch, dass es neben der passiven Demethylierung andere aktive Mechanismen zur Entfernung der Methylgruppe geben muss. In diesem Zusammenhang war die Entdeckung von weiteren DNA-Modifikationen, die durch enzymatische Oxidation von mdC entstehen ein großer Hoffnungsschimmer im Hinblick auf die weitere Aufklärung von Mechanismen der aktiven Demethylierung. Die neu entdeckten mdC Modifikationen besitzen durch die oxidierte Methylgruppe eine andere Reaktivität im Vergleich zu mdC und könnten an einem neuartigen, aktiven Demethylierungsmechanismus beteiligt sein. Im folgenden Abschnitt wird näher auf diese oxidierten mdC Modifikationen, deren Entdeckung, den Mechanismus deren Entstehung, Reaktivität und mögliche epigenetische Relevanz eingegangen.

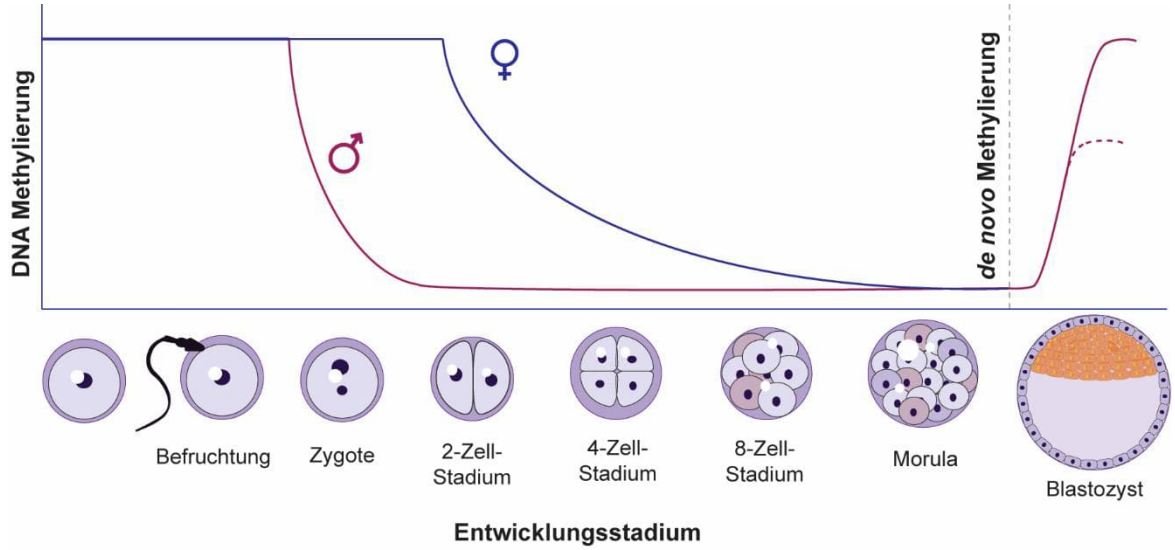


Abbildung 1.6: Entwicklung der globalen mdC Level nach der Befruchtung der Eizelle bis hin zur Ausbildung eines Blastozysts. Kurz nach der Befruchtung der Eizelle mit dem Spermium kommt es im paternalen Genom durch aktive Demethylierung zu einem drastischen Abfall der globalen mdC Level. Während der mdC Gehalt des paternalen Genoms für mehrere Zellteilungsschritte niedrig bleibt, kommt es im maternalen Genom zu einem schrittweisen langsamen Abfall der DNA Methylierung. Bei der Entwicklung des Blastozysts werden durch die *de novo* Methyltransferasen Dnmt3a und 3b die Methylierungsmuster eingeführt. Abbildung frei berarbeitet nach Wu *et al*.[8]

1.2.3 Nicht-kanonische DNA-Modifikationen hmdC, fdC und cadC

Die häufigste DNA Modifikation, mdC, wurde jahrzehntelang als die ”fünfte Base des Genoms” betrachtet und als eine der wichtigsten epigenetisch relevanten DNA Modifikationen wahrgenommen.[59] Nach der Entdeckung von 5-Hydroxymethyl-2’-desoxycytidin (hmdC), eines Oxidationsprodukts von mdC, wurde diese Modifikation anfangs nur als potentielle DNA Läsion wahrgenommen.[60] Im Jahr 2009 wurde in zwei Schlüsselstudien von *Heintz*[61] und

Rao[62] unabhängig voneinander hmdC in genomischer DNA von Säugetieren nachgewiesen. Es konnte zudem gezeigt werden, dass diese Modifikation eine wichtige Rolle in der epigenetischen Kontrolle von Neuronen[61] und embryonalen Stammzellen[62] spielt. Die Studie von *Rao* konnte weiterhin zeigen, dass die α-Ketoglutarat (α-KG) abhängige Fe(II)-dioxygenase Tet1 (*ten eleven translocation*) durch kontrollierte Oxidation von mdC für die Entstehung von hmdC verantwortlich ist.[62] Die Menge von hmdC im Genom ist in hohem Maße vom Organismus und Typ des Gewebes abhängig und kann stark schwanken. In verschiedenen Mausorganen lässt sich hmdC im Bereich von 0.05% bis 0.15% aller dCs quantifizieren[63-65], wobei im Gehirn deutlich höhere Mengen bis zu 1.5% nachweisbar sind.[61, 66] Auch im Gehirn ist jedoch die hmdC Konzentration sehr stark von der untersuchten Gehirnregion abhängig und akkumuliert mit steigendem Alter.[67-68] Ein weiterer wichtiger Durchbruch in diesem Feld wurde 2011 mit der Entdeckung der weiteren oxidierten mdC Modifikationen 5-Formyl-2'-desoxycytidin[69] (fdC) und 5-Carboxyl-2'-desoxycytidin[70-71] (cadC) erzielt. Diese biosynthetisch durch Tet katalysierte Oxidation von mdC im Genom entstehenden Modifikationen wurden im Genom von embryonalen Stammzellen der Maus in signifikanten Mengen detektiert.[35, 70] Analog zur detektierten Menge von hmdC im Genom, ist die Häufigkeit von fdC und cadC stark gewebeabhängig. Für das aus einer Tet-vermittelten Oxidation von hmdC resultierende fdC kann eine genomische Häufigkeit von ca. 0.02% pro dC ermittelt werden.[35, 59] Die höher oxidierte Spezies, cadC, kommt im Genom mit ca. 0.002% pro dC etwa 10-Mal mal seltener vor.[35, 59, 72]

Alle oxidierten mdC Derivate entstehen im Genom durch einen katalytischen Prozess ausgelöst durch eisenhaltige Oxygenasen, den sogenannten Tet-Enzymen. Der genaue Mechanismus der α-KG abhängigen Tet(1-3) vermittelten Oxidation von mdC zu hmdC, hmdC zu fdC und fdC zu cadC ist weitgehend aufgeklärt.[56, 58] Die wichtigsten Merkmale dieses Oxidationsprozesses wurden basierend auf zahlreichen Röntgen-diffraktometrie Studien des aktiven Zentrums der Tet-Enzyme postuliert.[73-74] Die Schritte sind am Beispiel der Tet1 katalysierten Hydroxymethylierung von mdC in **Abbildung 1.7** abgebildet.[75-78] In der Koordinationssphäre des Fe(II)-aktiven Zentrums **A** werden zwei Aqua-Liganden durch α-KG verdrängt welches dann zwei Koordinationsstellen unter Ausbildung des Fe(II)-Komplexes **B** okkupiert. Anschließend erfolgt der Schlüsselschritt der C-H Oxidation, eine Vier-Elektronen-Reduktion des Sauerstoffs unter Ausbildung der cyclischen Fe(IV)-Peroxo Spezies **C**.[79-81] Im nächsten Schritt wird durch mehrere *Elekronenshifts* der Fe(IV)-Oxo Komplex **D** ausgebildet, während α-KG unter Freisetzung von CO_2 zu Succinat umgewandelt wird.[81-84] Daraufhin kommt es über einen radikalischen Mechanismus zur Übertragung des Wasserstoffatoms der C5-Methylgruppe von mdC auf die Fe(IV)-Oxo Spezies **E**. Das Fe-Zentrum wird hierbei von Fe(IV) zu Fe(III) unter Ausbildung der Fe(III)-Hydroxyspezies **F** reduziert. Das C-H Radikal am

mdC wird anschließend über einen weiteren radikalischen Schritt mit der Fe(III)-Hydroxyspezies **F** unter Ausbildung von hmdC oxidiert und das aktive Zentrum wird in einer Ein-Elektronen-Reduktion zu Fe(II) **G** reduziert.[81] Ein Ligandentausch und erneute Koordination von zwei Aqua-Liganden führt zur Regeneration des aktiven Fe(II)-Komplexes **A**, der für einen neuen katalytischen Zyklus zur Verfügung steht.

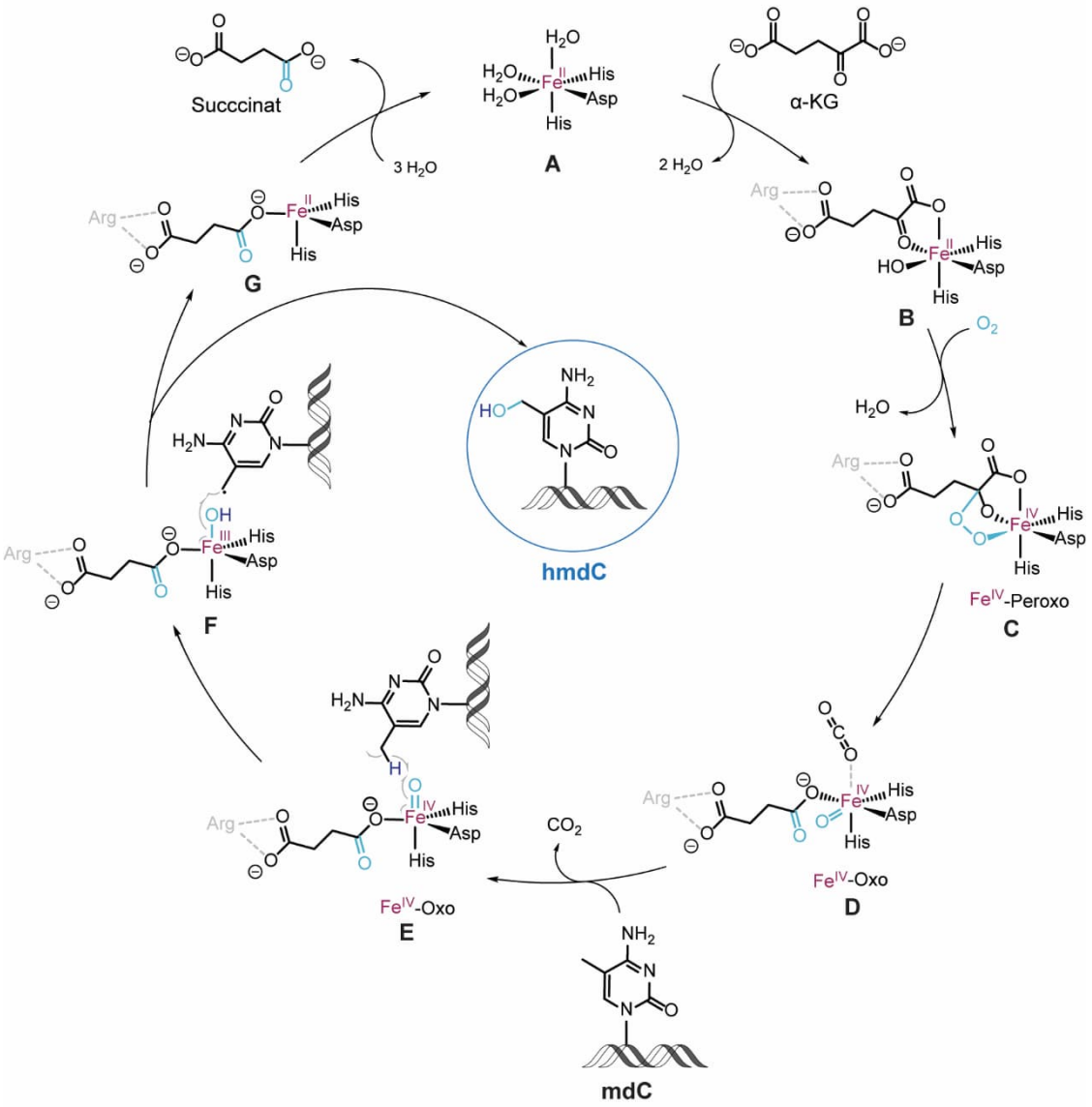


Abbildung 1.7: Postulierter Mechanismus der Tet-vermittelten Oxidation von mdC zu hmdC. Das Fe(II)-haltige aktive Zentrum des Enzyms **A** wird durch Koordination von α-KG und molekularem Sauerstoff in die aktive Fe-(IV)-Peroxo-Spezies **C** umgewandelt. Eine Decarboxylierung von α-KG zu Succinat liefert den Fe-(IV)-Oxo-Komplex **D**, welcher in einem radikalischen Schritt ein Wasserstoffatom von mdC abstrahiert. Eine anschließende Oxidation des mdC-Radikals durch die Fe-(III)-Hydroxospezies **F** führt zu hmdC und nach Freisetzung von Succinat zur Regeneration des aktiven Fe-(II) Zentrums **A**.

1.3 Aktive DNA Demethylierung – fdC und cadC als Intermediate der epigenetischen Regulation

Die passive Demethylierung, bei der es durch mehrere Zellteilungsschritte bei gleichzeitiger Deaktivierung von Dnmt1 zur Verdünnung von mdC im Genom kommt ist stark replikationsabhängig und kann nicht die ganze Bandbreite der Demethylierungsdynamik erklären. Die rapide Abnahme der mdC Menge im väterlichen Genom der Zygote nach der Befruchtung (**Abbildung 1.6**), die im Rahmen einer Zellteilung auftritt, kann nicht allein durch einen passiven Mechanismus erklärt werden.[85] Durch die Entdeckung von hmdC wurde die durch Tet-Oxidation gebildete Modifikation als mögliches Intermediat eines aktiven Demethylierungsmechanismus diskutiert.[58] Außerdem wurde gezeigt, dass, obwohl die Mengen an mdC im Entwicklungsstadium der Zelle rasch abnehmen, eine Zunahme der hmdC Menge zu beobachten ist. Dies stellt einen wichtigen Hinweis im Hinblick auf die Rolle der mdC Tet-Oxidationsprodukte bei der Zellreprogrammierung dar.[86] Wie bereits im vorangegangenen Kapitel beschrieben, sind die Mengen der oxidierten mdC Modifikationen hmdC, fdC und cadC stark gewebeanhängig. Generell sinkt die genomische Häufigkeit der Modifikation mit steigender Oxidationsstufe des C-5 Kohlenstoffs der Nukleobase.[35, 59, 72] In genomischer DNA kann fdC ungefähr 0.02% aller dCs darstellen, wobei die Häufigkeit von cadC mit 0.002% aller dCs noch 10-Mal seltener ist.[35, 59, 72] Diese höher oxidierten mdC Modifikationen sind somit nur in sehr kleinen Mengen im Genom vorhanden und wurden ursprünglich für kurzlebige Intermediate gehalten. Mit Hilfe von Verfütterungsexperimenten von isotopenmarkiertem Material konnte eindeutig gezeigt werden, dass genomisches fdC *in vivo* langlebig ist und positionsabhängig auch eine semi-stabile Modifikation des Genoms sein kann.[87-88] Die Schlussfolgerung aus der genomischen Stabilität von fdC war die potenzielle funktionelle Rolle dieser Modifikation als epigenetischer Marker und deren Beteiligung an der aktiven Demethylierung von mdC. Durch die hohe örtliche sowie zeitaufgelöste Spezifizität der DNA-Demethylierung war schon vor der Entdeckung der höher oxidierten mdC Modifikationen klar, dass es neben der replikationsabhängigen passiven DNA-Demethylierung einen anderen, aktiven Mechanismus geben muss.[8] Ein direkter Bruch der sehr stabilen C-C-Bindung von mdC ist eine thermodynamisch extrem energieaufwändige Reaktion die so ohne weiteres von einem Enzym nicht durchgeführt werden kann. In der ersten Studie, die eine direkte Entfernung der Methylgruppe von mdC durch enzymatische Beteiligung postuliert hatte, wurde das Protein MBD2 (*Methyl-CpG binding domain protein 2*) als potentielle mdC Demethylase vorgeschlagen.[89-90] Diese kontroverse Studie konnte jedoch nicht erklären, wie Organismen mit einem MBD2 Defizit weiterhin normale Demethylierungsmuster aufwiesen.[91-92] Nachdem die Ergebnisse von keiner anderen Forschungsgruppe reproduziert werden konnten, entstanden große Zweifel an der Beteiligung von MBD2 an der direkten Demethylierung von

mdC. In diesem Zusammenhang zeigte die Entdeckung von fdC und cadC neue richtungsweisende Perspektiven für die mechanistischen Überlegungen zur aktiven Demethylierung von mdC auf. Zum einen können die Modifikationen fdC und cadC durch Glykosylasen über den Basenexzisionsreparaturweg demodifiziert werden und dadurch zur aktiven DNA-Demethylierung beitragen. Zum anderen ist bei fdC und cadC durch die erhöhte Polarität und Reaktivität der exozyklischen C-C-Bindung im Vergleich zu mdC ein direkter C-C-Bindungsbruch aus chemischer Sicht vorstellbar. In **Abbildung 1.8** sind mögliche chemische Transformationen zusammengefasst, die auf der Genomebene stattfinden können um mdC in dC umzuwandeln. In den nächsten Kapiteln werden diese Transformationen und Mechanismen der replikationsunabhängigen aktiven DNA-Demethylierung, bei denen fdC und cadC eine zentrale Rolle spielen, detailliert diskutiert.

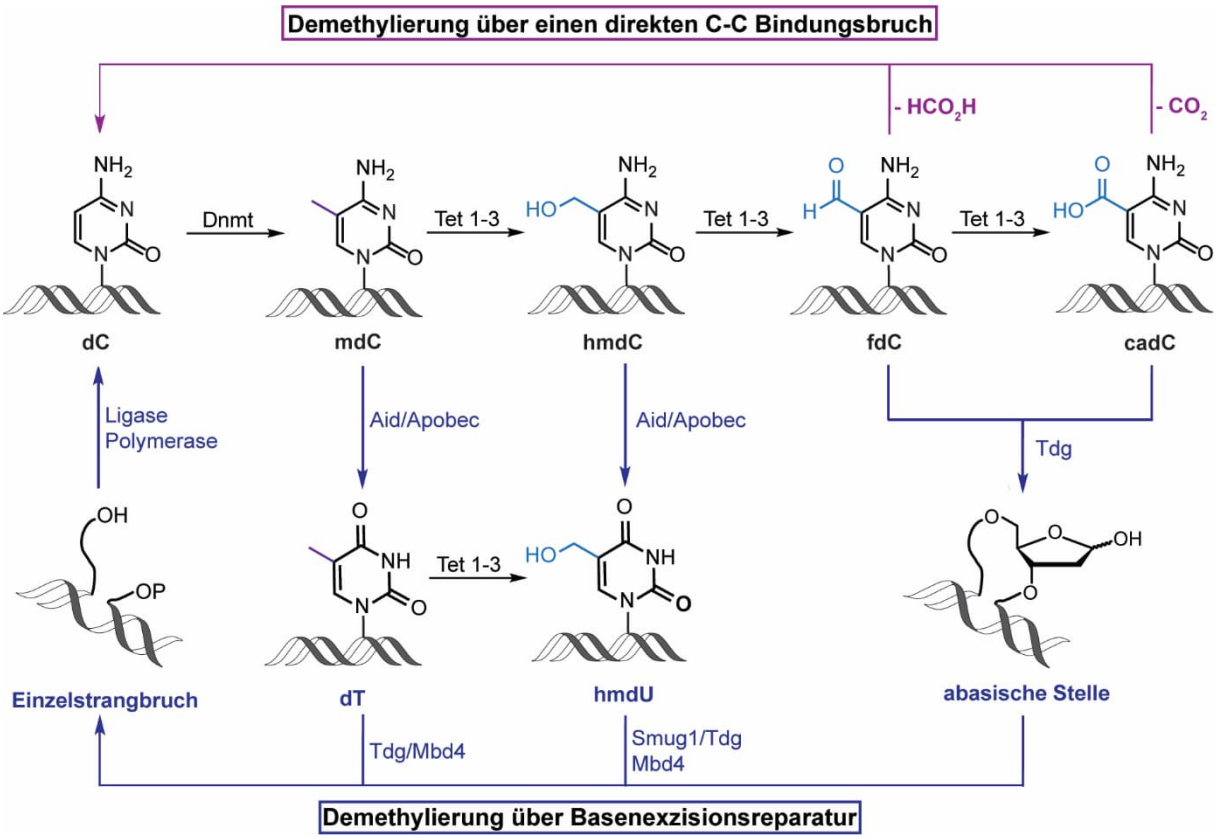


Abbildung 1.8: Zusammenfassung von unterschiedlichen chemischen Wegen der aktiven DNA Demethylierung. Genomisches dC wird durch Dnmts zu mdC methyliert. Die methylierte Modifikation kann iterativ durch Tet-Enzyme zu hmdC, fdC und cadC oxidiert werden. Im oberen Teil der Abbildung (dargestellt in lila) ist die Demethylierung über einen direkten C-C Bindngsbruch dargestellt. Die oxidierten Nukleobasen fdC und cadC können durch eine Deformylierungs- (Freisetzung von Ameisensäure) bzw. Decarboxylierungsreaktion (Freisetzung von CO_2) direkt zu dC demodifiziert werden. Der untere Teil der Abbildung (dargestellt in blau) beschreibt die Entfernung des mdC aus dem Genom über Basenexzisionsreparatur. Dies kann über eine direkte Entfernung der fdC bzw. cadC Nukleobase mit Hilfe von der Tdg unter Ausbilung einer abasischen Stelle erfolgen, welche im weiteren Verlauf von BER prozessiert und mit kanonischem dC ersetzt wird. Außerdem kann es zu einer aktiven Demethylierung kommen, indem mdC und hmdC in einer Desaminierungsreaktion zu dT und hmdU umgewandelt werden, welche im Rahmen der BER weiter prozessiert werden.

1.3.1 Aktive DNA Demethylierung über Basenexzisionsreparatur von oxidierten mdC-Modifikationen

Die Entdeckung der oxidierten mdC-Modifikationen eröffnete neue Mechanismen für das Verständnis der epigenetischen Regulationsprozesse auf DNA-Ebene. Ein etablierter Mechanismus für die Entfernung von mdC aus dem Genom ist die Prozessierung von fdC und cadC im Rahmen der Basenexzissionsreparatur (BER). Die oxidierten mdC-Modifikationen fdC und cadC stellen daher Substrate für die BER dar und können unter Erzeugung abasischer Stellen aus dem Genom entfernt werden.[71, 93-94] Dies geschieht durch die Thymin DNA Glykosylase (Tdg), die für die Erkennung und Reparatur von Guanin:Thymin Fehlpaarungen zuständig ist.[95] Neben der dG:dT Reparatur ist Tdg in der Lage selektiv fdC und cadC im Genom zu erkennen und deren glykosidische Bindung zu spalten.[71, 95-96] Bei der Entfernung von Basen aus dem Genom im Zuge der BER wird im ersten Schritt die nicht-kanonische oder geschädigte Base von einer Glykosylase erkannt und im Anschluss wird die Bindung zwischen der Ribose und dem DNA-Heterozyklus gespalten, wobei eine abasische Stelle entsteht.[97-98] Bei den Glykosylasen wird je nach Ablauf von BER zwischen mono- und bifunktionalen Enzymen unterschieden. Monofunktionale Glykosylasen brechen die *N*-glykosidische Bindung des Nukleosids unter Entstehung einer abasischen Stelle, welche im Folgeschritt über Endonukleasen, wie z. B. die *apurinic/apyrimidinic endonuclease* (Ape1) zu einem 5'-Phosphatester und einem freien 3'-OH prozessiert wird.[99] Bifunktionale Glykosylasen sind in der Lage neben der initialen Spaltung der *N*-glykosidischen Bindung der Nukleobase durch ihre Lyasefunktion eine weitere Prozessierung der abasischen Stelle durchzuführen. Hierbei erfolgt eine Abstraktion des 2'-Protons der abasischen Stelle. Im Rahmen einer β-Eliminierung bildet sich dann eine 2'-3'-ungesättigte Bindung. Eine anschließende δ-Eliminierungsreaktion führt dann unter Freisetzung von 4-Oxopent-2-enal zu einem DNA Strangbruch.[100-102] Der Strangbruch kann durch den Einbau eines kanonischen Nukleosids und Ligation durch eine Polymerase repariert werden.[100] Eine schematische Darstellung der BER ist in **Abbildung 1.9** dargestellt.

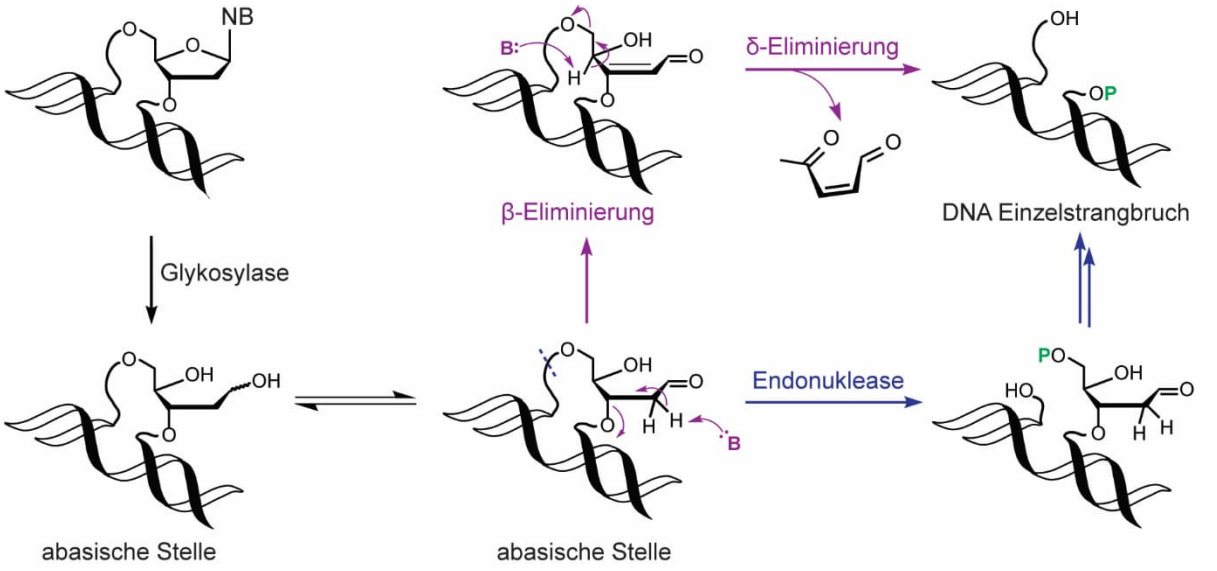


Abbildung 1.9: Ablauf der Basenexzisionsreparatur initiiert durch die Spaltung der *N*-glykosidischen Bindung des Nukleosids durch die Aktivität einer Glykosylase unter Entstehung einer abasischen Stelle. Die abasische Stelle kann weiter entweder durch die Aktivität einer Endonuklease oder durch eine bifunktionalen Glykosylase über eine β- und δ-Eliminierung weiterprozessiert werden.

Ein weiterer postulierter aktiver Demethylierungsweg von mdC verläuft über eine enzymatische Desaminierungsreaktion von mdC zu dT mit anschließender BER. Genomisches mdC kann durch den Cytidindesaminasekomplex (Aid/Apobec)[8, 94] oder Dnmts[103] zu dT umgewandelt werden, wobei eine dG:dT Fehlpaarung entsteht.[103] Ein derartiger *mismatch* wird von Thymidin-Glykosylasen wie Tdg und MBD4 erkannt und mittels Exzision fehlerhafter Basen repariert.[8, 104-109] Es gibt Hinweise auf die Beteiligung einer enzymatischen Desaminierungsreaktion im Rahmen der aktiven mdC Demethylierung, jedoch zeigten Studien mit Desaminase-defizienten Mäusen (*deaminase-enzyme-knockout mice*), dass ein Fehlen des Enzyms zu keinen Defekten in der Entwicklung und Reproduktion führte.[110-111] Auch ist zu erwähnen, dass die Aid/Apobec-Proteine unmodifizierte dCs, sowie einzelsträngige DNA bervorzugen, was deren Beteiligung an der aktiven Demethylierung in Frage stellt.[112] Die Desaminierungsaktivität von Dnmts wird allerdings eher kritisch betrachtet, da diese nur bei einer sehr geringen SAM Konzentrationen auftritt.[103] In Anbetracht der essentiellen Rolle von SAM in zahlreichen biochemischen Vorgängen und dessen Rolle bei der Dnmt katalysierten Methylierung von DNA ist eine sehr geringe und hoch fluktuierende SAM Konzentration eher unwahrscheinlich.[8] Das Desaminasesystem Aid/Apobec führt außerdem zur Desaminierung von hmdC zu hmdU. Hierbei entsteht eine hmdU:dG Fehlpaarung, welche auch ein Substrat für BER darstellt und über Tdg/Mbd4 ausgeschnitten werden kann.[94, 113-114] Isotopenverfolgungsexperimente mit markierten Nukleosiden konnten zeigen, dass die Desaminierung von hmdC nicht zum Erreichen von *steady-state* Level von hmdU in mESCs beiträgt, wodurch dieser desaminierungsvermittelte, an BER gekoppelte DNA

Demethylierungsweg biologisch wenig relevant erscheint.[115] Neben der Desaminierung kann hmdU im Genom von Stammzellen auch über eine Tet-vermittelte Oxidation von dT auftreten.[115] Dieser Prozess kann zu einer hmdU:dA Fehlpaarung führen, und durch die spezielle hmdU-Glykosylase Smug1 (*Single-strand selective monofunctional uracil DNA glycosylase*) prozessiert werden.[94, 116] Im Fall der DNA Demethylierung von fdC und cadC über die BER wird der entstehende Strangbruch durch die Aktivität einer Polymerase und Ligase repariert, indem ein kanonisches dC inkorporiert wird.[100] Dieser Demethylierungsprozess verläuft über abasische Stellen, die wie bereits beschrieben zu Strangbrüchen führen können. mdC tritt im Genom häufig geclustert auf und dCs können symmetrisch auf gegenüberliegenden Stellen im Genom methyliert vorliegen. Dadurch kann es potentiell durch die BER zu schädlichen Doppelstrangrüchen kommen[117], die zu Apoptose[118], Arrest des Zellzyklus[118] sowie DNA Mutationen[119] führen können. In Systemen mit einer induzierten reduzierten Tdg-Aktivität (Tdg-*Knockout*) konnte eine 10-fache Akkumulation von fdC und cadC beobachtet werden.[71, 88, 120-122] Eine Studie, bei der potentiell schädliche abasische Stellen intermediär abgefangen und quantifiziert wurden, konnte jedoch zeigen, dass BER von oxidierten mdC Modifikationen zu keiner Akkumulation von schädlichen Intermediaten im Genom führt.[123] Daraus lässt sich schlussfolgern, dass die Tdg-vermittelte DNA-Demethylierung entweder keine vorrangige Rolle in höheren Organismen einnimmt oder hochorchestriert über einen komplexen Multienzymkomplex abläuft, welcher die abasischen Stellen sofort weiter prozessiert und mit dem entspechenden Nukleosid füllt.[124-125] Neil-Proteine sind Enzyme, welche in ihrer Rolle als bifunktionale Glykosylasen für die Entfernung der oxidativen Schäden wie 5-Hydroxy-cytosin und 8-Oxo-dG verantwortlich sind.[126-128] Diese wurden auch als fdC-Binder und somit als ein Teil der Tdg vermittelten Demethylierung identifiziert.[129] Eine weitere Studie zeigte in diesem Zusammenhang, dass Neil 1 und 2 für die Prozessierung von fdC und cadC während BER essentiell sind.[125] Da es sich bei Tdg um eine produktinhibierte Glykosylase handelt, ist ein weiterer Schritt erforderlich, um das Enzym vom DNA-Strang freizusetzen. Die Proteine Neil 1 und 2 setzen nach der Spaltung der *N*-glykosidischen Bindung Tdg vermutlich von der abasischen Stelle frei, welches dort kovalent gebunden ist.[130] Die Tdg-vermittelte BER von fdC und cadC ist ein gut untersuchter und akzeptierter aktiver Demethylierungsweg.[8, 59] Die damit verbundene Bildung von potentiell schädlichen DNA Strangbrüchen, sowie der Fakt, dass in Tdg-defizienten Zygoten eine Tet3-vermittelte, aber Tdg-unabhängige aktive Demethylierung abläuft, waren eindeutige Hinweise für einen weiteren aktiven Demethylierungsmechanismus.[131] Ein aus chemischer Sicht attraktiver komplementärer Mechanismus verläuft über einen direkten Bruch der C-C-Bindung an der C-5 Position der Nukleobase von mdC oder dessen oxidierter Derivate. Im folgenden Abschnitt wird dieser direkte DNA-Demethylierungsweg geschildert und im Detail erläutert.

1.3.2 Aktive DNA Demethylierung über einen direkten C-C-Bindungsbruch

Neben der Tdg-abhängigen DNA Demethylierung über BER wurde in den letzten Jahren ein weiterer Demethylierungsmechanismus, welcher über einen direkten C-C-Bindungsbruch ablaufen könnte, intensiv diskutiert.[8, 65] Die Idee einer direkten Entfernung der Methylgruppe aus mdC wurde schon 1999 im Zusammenhang mit der Untersuchung des Mbd2 Proteins postuliert. Dieses Enzym wurde als eine potentielle Demethylase beschrieben, welches mdC unter Abspaltung von Methanol zu dC umsetzt.[89-90] Eine derartige direkte Spaltung der C-C-Bindung zwischen dem sp^3-Kohlenstoff der Methylgruppe und einem aromatischen sp^2-Kohlenstoff der Nukleobase stellt eine thermodynamisch extrem energieaufwändige Reaktion dar. Eine Katalyse dieser Reaktion durch Mbd2 konnte in weiteren Studien nicht bestätigt werden. Im Weiteren konnte gezeigt werden, dass die DNA-Methyltransferasen Dnmt1, 3a und 3b in Abwesenheit von SAM mdC unter C-C-Bindungsspaltung direkt zu dC demethylieren können.[132] Dieser Demethylierungsmechanismus basiert auf der Abwesenheit von SAM und erscheint aufgrund dessen hoher intrazellulärer Konzentration biologisch irrelevant. Studien zu chemischen Eigenschaften und zur Reaktivität der mdC-Oxidationsprodukte hmdC, fdC und cadC zeigten, dass neben der steigenden Polarität der C-C-Bindung, auch die Reaktivität der Modifikation mit zunehmendem Oxidationsgrad ansteigt.[93, 133] Kurz nach der Entdeckung von fdC wurde ein Demethylierungsmechanimsus postuliert, der über einen direkten C-C-Bindungsbruch an der C-5 Position der formylierten Nukleobase abläuft.[35, 134] In *in vitro* Studien mit fdC und cadC wurde gezeigt, dass diese Modifikationen durch Zugabe von Nukleophilen wie Thiolen effizient zu dC defunktionalisieren.[93, 133] Interessanterweise verläuft die Decaboxylierung von cadC mit einer 11-fach höheren Reaktionsrate als die analoge Demodifizierung von fdC.[133] Modellstudien und quantenchemische *ab initio* Rechnungen zeigten, dass die Deformylierungs- und Decarboxylierungsreaktion eine Aktivierung der Nukleobase über einen nukleophilen Angriff an die C-6 Position des Cytosinrings benötigt. In einer derart aktivierten Nukleobase kann es unter Freisetzung des Nukleophils zur Demodifizierung mittels eines C-C-Bindungsbruchs an C-5 kommen.[133] *In vitro* Studien konnten in diesem Zusammenhang durch Einsatz von synthetischen, cadC-enthaltenden Oligonukleotiden eine derartige Demodifizierung nach Inkubation mit Stammzellextrakten zeigen.[93] Mechanistisch ähnelt der postulierte Ablauf des C-C-Bindungsbruchs der Dnmt-katalysierten Methylierung von dC mit SAM als Kofaktor. Bei der Methylierung von dC kommt es nach der säurekatalysierten Sättigung der C5-C6-Bindung der Nukleobase durch Angriff eines Nukleophils zur Funktionalisierung an C-5 mit anschließender Rearomatisierung (**Abbildung 1.4**). Für die Deformylierungs- bzw. Decarboxylierungsreaktion von fdC bzw. cadC wurde analog eine Aktivierung durch einen nukleophilen Angriff einer Aminosäureseitenkette wie Cystein an der C-6 Position des Pyrimidinrings postuliert.

Experimentelle Unterstützung für den postulierten Mechanismus lieferte eine erfolgreiche Defunktionalisierung der Nukleoside nach Umsetzung mit Cystein und basischen Aminosäuren, die das aktive Zentrum eines potentiellen Enzyms simulieren sollten.[35, 93] In **Abbildung 1.10** ist der mechanistische Ablauf einer Deformylierungsreaktion von fdC bzw. einer Decarboxylierung von cadC mit Cystein als Helfernukleophil postuliert. Ein genomisches fdC kann z. B. durch eine nukleophile Cysteinseitenkette eines potenziellen Enzyms Michael-artig an der C-6 Position angegriffen werden. Über eine Enamin-Imin-Tautomerisierung kommt es zur Sättigung der C-5-C-6-Bindung der Nukleobase. In wässriger Umgebung entsteht dann ein Gleichgewicht zwischen der Aldehyd- und der Hydratform der Carbonylfunktionalität von fdC. Eine Deprotonierung der exozyklischen Aldehydhydrateinheit des gesättigten fdC Intermediats führt unter Eliminierung des Helfernukleophils zu einer Abspaltung von Ameisensäure und damit schlussendlich zu einer Deformylierung von fdC zu dC. Für die Decarboxylierung von cadC kann analog die Aktivierung der carboxylierten Nukleobase durch einen nukleophilen Angriff eines Helfernukleophils erfolgen. Eine Enamin-Imin Tautomerie führt zur Sättigung der C-5-C-6-Bindung. Anschließend kann über einen sechsgliedrigen Übergangszustand unter Eliminierung des Helfernukleophils und Rearomatisierung der Nukleobase eine Decarboxylierung von cadC erfolgen. Umgebende Aminosäuren können bei den Demodifizierungsreaktionen als Protonen-Shuttle fungieren.

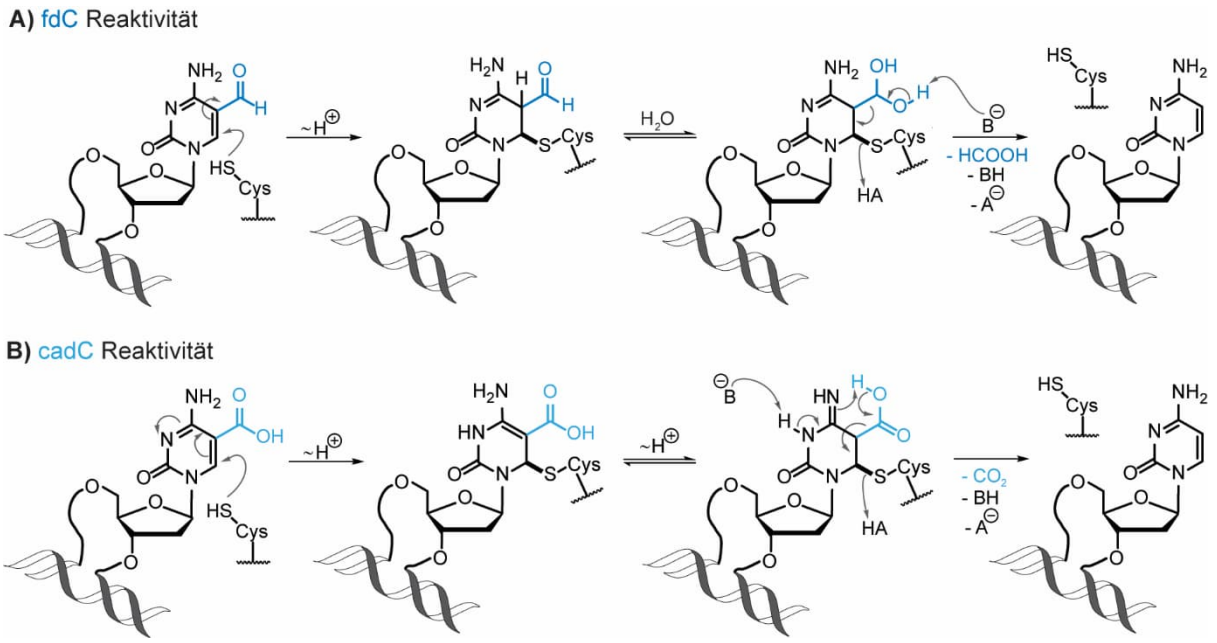


Abbildung 1.10: A) Postulierter Mechanismus einer fdC Deformylierung zu dC. Angriff eines Helfernukleophils und eine Enamin-Imin-Tautomerie führt zur Sättigung der C5-C-6-Bindung. Deprotonierung der Hydrat-Form von fdC in wässriger Umgebung kann unter Freisetzung von Ameisensäure und Eliminierung von Cystein unter Deformylierung zu dC führen. **B)** Postulierter Mechanismus der Decarboxylierung von cadC zu dC. Cystein führt zur Sättigung der C5-C6-Bindung der Nukleobase. Nach erfolgter Enamin-Imin-Tautomerie kann über einen sechsgliedrigen Übergangszustand unter Abspaltung des Helfernukleophils und Rearomatisierung eine Decarboxylierung zu dC stattfinden. (Helfernukleophil = z. B. Cystein, B = Base, HA = Säure).

Die erste experimentelle *in vivo* Deformylierungsstudie konnte 2017 beweisen, dass im Genom von mESCs die aktive Demethylierung unabhängig von Tdg über einen direkten C-C-Bindungsbruch in Form von Deformylierung von fdC abläuft.[135] Durch den Einsatz von chemisch modifizierten Nukleosidderivaten in zellbiologischen Verfütterungsexperimenten konnte zum ersten Mal eine genomische Deformylierung von fdC experimentell bestätigt werden (**Abbildung 1.11**). In den hierzu verwendeten Experimenten wurden chemisch modifizierte Sondenmoleküle (*probe molecules*) an Zellkulturen von mESCs verfüttert und als Reportermoleküle in deren Genom eingebaut. Eine anschließende Isolation der genomischen DNA, gefolgt von einem Totalverdau auf Nukleosidebene und Analyse mittels hochsensitiver Massen-spektrometrie gekoppelt an Ultrahochleistungs-Flüssigchromatographie konnte chemische Veränderungen auf Molekülebene verfolgen. So konnte beim Einsatz eines $^{13}C_5$-$^{15}N_2$-isotopenmarkierten fdC Reportermoleküls das Deformylierungsprodukt $^{13}C_5$-$^{15}N_2$-dC im Genom der untersuchten Zellen detektiert werden. Neben isotopenmarkierten Testmolekülen wurden auch 2'-fluorierte Nukleosidderivate eingesetzt, die aufgrund ihrer veränderten elektronischen Eigenschaften der Ribose nicht von Tdg ausgeschnitten und über BER entfernt werden können. Diese ribosemodifizierten Sondenmoleküle konnten so als ideale Modellsysteme für die gezielte Tdg-entkoppelte Untersuchung der Demethylierung über einen direkten C-C-Bindungsbruch dienen. Bei Zellverfütterungsexperimenten von 2'-F-fdC konnte *in vivo* eine hohe genomische Deformylierungsaktivität zu 2'-F-dC nachgewiesen werden. Zur Bestätigung, dass es sich bei der Deformylierungsreaktion um einen genomischen Prozess handelt, wurde neben der genomischen DNA auch der lösliche Nukleosidpool der Zellen untersucht. Da hier kein potenzielles Deformylierungsprodukt detektiert wurde, konnte angenommen werden, dass die Deformylierungsreaktion im Genom stattfindet.

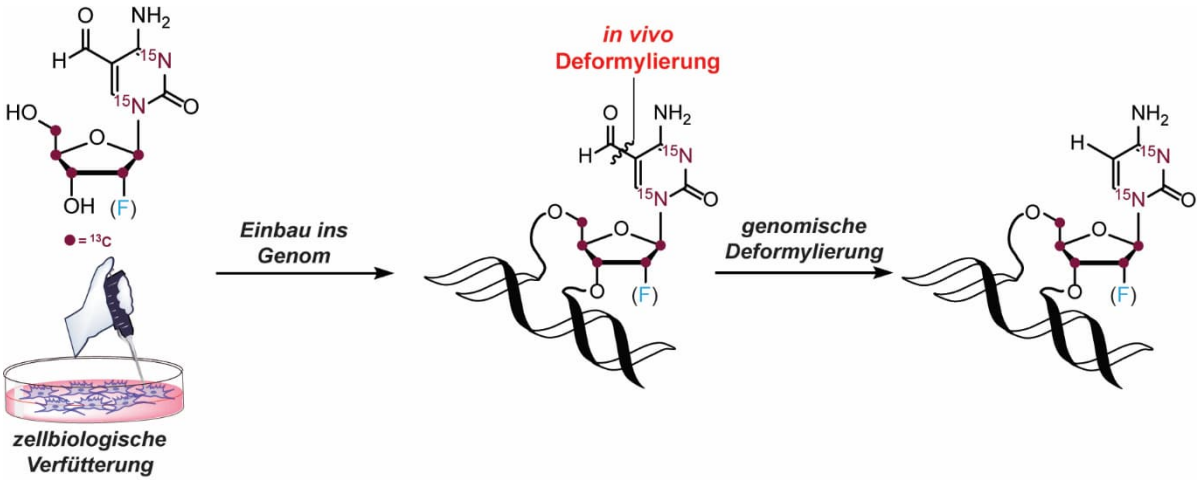


Abbildung 1.11: Experimentelle Herangehensweise der ersten *in vivo* Deformylierungsstudie. Isotopenmarkierte und 2'-fluorierte Nukleosidderivate wurden an Stammzellkukturen verfüttert und in deren Genome eingebaut. Im Genom konnte eine Umsetzung zum entsprechenden deformylierten dC-Derivat beobachtet werden.[135]

Die Frage, ob dieser Demethylierungsmechanismus in einer großen Bandbreite von Zellsystemen und Organismen ausgeprägt ist und ob eine und wenn ja welche enzymatische Entität für die Deformylierung zuständig ist, bleibt bis heute ungeklärt. Die Rolle von cadC in der aktiven DNA-Demethylierung über eine Decarboxylierung von cadC wurde in den letzten Jahren auch zunehmend intensiv untersucht. Eine Decarboxylierung von cadC konnte bereits kurz nach der Entdeckung dieser Modifikation *in vitro* in mehreren Studien gezeigt werden. Es wurde sowohl eine Decarboxylierungsaktivität von cadC durch den Einsatz von Stammzelllysaten[93, 136] als auch durch Dnmt3a und 3b in Abwesenheit von SAM[137] beobachtet. Der Ausschluss von SAM macht jedoch diese Beobachtung *in vivo* irrelevant. Es bleibt eine offene Frage, ob cadC auch *in vivo* auf genomischer Ebene zu dC decarboxyliert und auf diese Weise an der aktiven DNA Demethylierung beteiligt ist. Aus dem *thymidine salvage pathway* (Wiederverwertung von Thymin) in dem Pilz *Neurospora crassa* ist eine vergleichbare enzymatisch katalysierte Umsetzung bekannt.[138] Dieser Pilz besitzt das Enzym Isoorotat Decarboxylase (IDcase), welche die Decarboxylierung von Isoorotat zu Uracil katalysiert. Die IDcase zeigte *in vitro* zusätzlich eine geringe Decarboxylierungsaktivität bei der Umsetzung von cadC zu dC.[139] Die Decarboxylierungsaktivität der IDcase war der erste enzymatische *in vitro* Beweis der Decarboxylierung von cadC zu dC und zeigte damit eindeutig, dass dieser C-C-Bindungsbruch unter enzymatischer Katalyse ablaufen kann. Aus chemischer Sicht ähnelt der *thymidine salvage pathway*, eine Umsetzung von T zu U, der DNA-Demethylierung, der Umsetzung von mdC zu dC.[8, 70, 93] Bei der in Pilzen ablaufenden Demethylierung von T zu U handelt es sich um einen aus mehreren Schritten bestehenden Prozess, der eine schrittweise Oxidation von T beinhaltet. Katalysiert durch Thymin-7-Hydroxylase (T7H) wird T in drei aufeinanderfolgenden Oxidationsschirtten zu hmU, fU und caU oxidiert. Die Demethylierung von T zu U wird über die IDcase katalysierte Decarboxylierung von caU erreicht.[140-144] In Analogie verläuft in Säugetieren auf DNA-Ebene die Tet-katalysierte Oxidation von mdC zu hmdC, fdC und cadC. Diese Analogien, sowie die IDcase cadC Decarboxylierungsaktivität legen die Vermutung nahe, dass eine potenzielle DNA-Decarboxylase Ähnlichkeiten in Sequenz, Struktur und katalytischem Mechanismus mit der IDcase aufweisen könnte. Das aktive Zentrum der IDcase besteht aus einem divalenten Metallion (mit hoher Wahrscheinlichkeit Zn^{2+}) und vier konservierten Aminosäureresten (Asp und His), die für die katalytische Decarboxylierung essentiell sind.[139] Beim Austausch der Carbonsäure-Funktionalität von caU mit einer Nitro-Funktionalität, die mit ihren stereoelektronischen Eingenschaften ein Analogon der Carbonsäure darstellt, wurde das 5-Nitro-substituierte Uridin als ein IDcase Inhibitor identifiziert.[139] Falls eine genomische Decarboxylierung von cadC enzymatisch abläuft und ein Mechanismus vergleichbar zur

Decarboxylierung der IDcase vorliegt, könnte ein 5-Nitro-substituiertes dC Derivat als ein potentieller Decarboxylase-Inhibitor fungieren.

Basierend auf den Studien der letzten Jahre zur genomischen Reaktivität von fdC und cadC wird deren Demodifizierung als ein Teil des aktiven Demethylierungsprozesses diskutiert.[93, 133, 135] Es ist *bis dato* nicht bekannt, an welchen Stellen im Genom bzw. zu welchem Zeitpunkt diese Prozesse ablaufen und ob es ein oder mehrere Enzyme gibt, die für diese Reaktivität zuständig sind. Weiterhin sind auch der genaue mechanistische Ablauf und die Struktur der potenziell involvierten Enzyme bis heute unbekannt.

1.4 Nicht-kanonische DNA und RNA Modifikationen

Neben den kanonischen Nukleobasen der DNA und RNA wurde eine Vielzahl an nicht-kanonischen modifizierten Nukleosiden identifiziert. Vor allem in der RNA sind heutzutage mehr als 160 posttranskriptional modifizierte Nukleobasen bekannt.[145-147] Durch den Einsatz von Flüssigchromatographie gekoppelt mit Massenspektrometrie konnten in den vergangenen Jahren stetig neue RNA-Modifikationen und deren biologische Funktion aufgeklärt werden.[145] Die hohe Anzahl an verschiedenen Modifikationen in der RNA korreliert mit den vielfältigen Funktionen der RNA in einem Organismus.[145-146] Die Komplexität der Modifikationen variert je nach Modifizierungsgrad und Ort des Vorkommens stark. Bezüglich ihrer Funktion spielen diese Modifikationen etwa eine wichtige Rolle zur Stabilisierung der mRNA[148-149], für Splicingprozesse[150-151], in tRNA und rRNA bei der Translation[152-153] oder im Rahmen der X-Chromosom Inaktivierung[154]. Chemische Modifikationen der Nukleoside können zum einen einen direkten Einfluss auf die RNA-Struktur haben und zu veränderten intramolekularen Interaktionen z. B. mit Proteinen führen. Zum anderen können strukturelle Veränderungen auf der RNA-Nukleosid Ebene zu einer funktionellen Veränderung des RNA Moleküls und dadurch zu Veränderungen im Zellmetabolismus führen.[145] Der Modifizierungsgrad der Nukleobasen kann von einfachen chemischen Funktionalisierungen wie Methylierung, Alkylierung, Thiomethylierung oder Acetylierung bis hin zu komplexeren Funktionalisierungen reichen. So treten zum Teil stark modifizierte Strukturen wie etwa Pseudouridin (Ψ)[155], hypermodifizierte 7-Deazaguanine[156-157] oder Pyrimidine auf. Ein strukturell besonders komplexes Beispiel einer nicht-kanonischen RNA Hypermodifikation stellt die Familie der Queuosine dar. Das Grundgerüst der Queuosine bildet eine 7-Deazaguanosin Einheit, die über eine Methylamin-Brücke mit einer Cyclopentendiol-Einheit verknüpft ist.[158] Queuosin (Q), das sich auf der Wobble-Position, Position 34 der tRNA, befindet, wurde bereits in 1972 entdeckt und strukturell aufgeklärt.[159] Später wurden weitere modifizierte Q-Derivate entdeckt, das biosynthetische Intermadiat Q-Epoxid (oQ),[160] die glykosylierten Varianten mit Mannose (ManQ) und

Galactose (GalQ),[157] und das aminosäuremodifizierte Glutamyl-Queuosin (GluQ).[161-162] Queosin befindet sich sowohl in bakterieller als auch eukaryotischer RNA, wobei in Eukaryoten keine *de novo* Biosynthese der Q-Base stattfindet. Stattdessen muss Queuosin aus der Nahrung aufgenommen werden und wird anschließend über eine Basenaustauschreaktion direkt in die tRNA eingebaut.[163] In Bakterien wird Q in zehn chemischen Transformationen aus GTP synthetisiert, wobei zwei der Biosyntheseschritte auf der tRNA Ebene ablaufen.[164-169] Bis heute ist die genaue Funktion der Modifikationen der Q-Familie nicht aufgeklärt.[158, 170] Es wurde eine Reihe an biologischen Erscheinungen beim Fehlen der Modifikation beobachtet, unter anderen neurologische Unregelmäßigkeiten[171], posttranslationale Proteinfehlfaltung[172] und Auffälligkeiten in Krebsgeweben.[172-174] Obwohl die glykosylierten Derivate ManQ und GalQ bereits 1976 entdeckt wurden und ein Strukturvorschlag für diese basierend auf NMR-Daten von isoliertem Material aus Hasenleber postuliert wurde, [157] sind die bisher vorliegenden analytischen Daten unzureichend und ein endgültiger Beweis der Struktur steht bis heute aus. In diesem Zusammenhang konnte 2020 von *Carell* und Mitarbeitern mittels Koinjektionsexperimenten von verdauter Maus-RNA mit einem synthetischen Standard gezeigt werden, dass die tatsächliche Struktur von GalQ mit der vor 35 Jahren postulierten Struktur übereinstimmt.[175] Für ManQ steht eine Verifizierung der Struktur noch aus. Bisherige Experimente zeigen jedoch, dass die postulierte Struktur von ManQ nicht mit der tatsächlichen Struktur übereinstimmt.[176] Daher sind weitere Experimente zur eindeutigen Aufklärung der Struktur dieser wichtigen RNA Modifikation von Interesse.

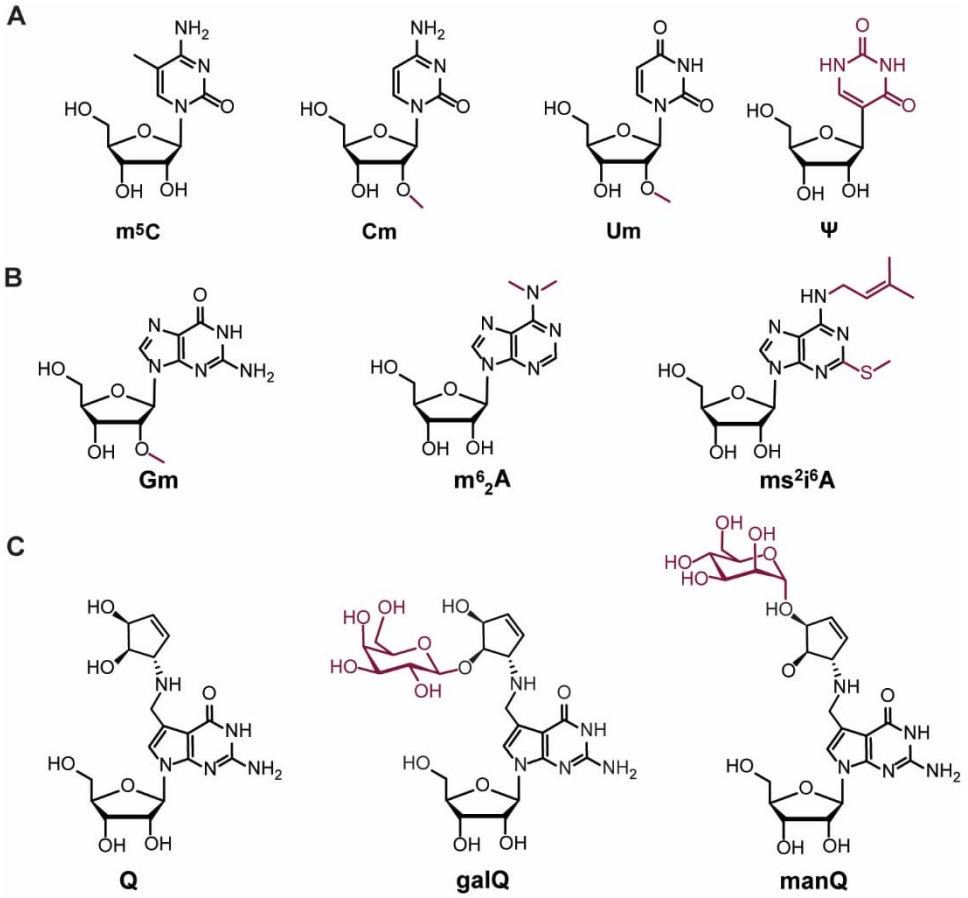


Abbildung 1.12: Eine Auswahl an nicht-kanonischen RNA Modifikationen**. A)** Einfache Modifikationen der Pyrimidinbase sowie eine Isomerisierung der Nukleobase zu Pseudouridin. **B)** Chemische Modifizierungen des Puringerüsts von A und G durch eine Alkylierung und Thiomethylierung. **C)** Familie der hypermodifizierten Queuosin-Modifikationen.

2 AUFGABENSTELLUNG

Der Fokus dieser Dissertation lag auf der Untersuchung der intragenomischen Reaktivität der oxidierten mdC Modifikationen fdC und cadC. Seit der Entdeckung dieser DNA-Modifikationen stellt sich die Frage, welche Funktion sie innerhalb des Genoms erfüllen und ob sie eine aktive Rolle in der DNA Demethylierung spielen. Nachdem *in vivo* gezeigt werden konnte, dass fdC in einer Deformylierungsreaktion innerhalb des Genoms zu kanonischem dC umgewandelt wird[135], stellte sich die Frage nach dem mechanistischen Ablauf dieser Reaktion sowie nach der potenziellen katalytischen Entität, welche für diesen Prozess zuständig sein könnte. Neben der Rolle von fdC im Demethylierungsprozess ist auch die Beteiligung von cadC nicht vollständig geklärt. Für cadC wurde eine chemisch attraktive genomische Decarboxylierungsreaktion zu dC postuliert, die einen Teil der aktiven Demethylierung darstellen könnte. Um tiefere Einblicke in den genomischen Deformylierungsprozess von fdC zu gewinnen und den genomischen Decarboxylierungsprozess von cadC *in vivo* zu untersuchen, sollten mit Hilfe von 2'-(*R*)-fluorierten und isotopenmarkierten Nukleosidderivaten Verfütterungsexperimente in einer Vielzahl von unterschiedlichen Zelllinien durchgeführt werden (**Abbildung 2.1**).

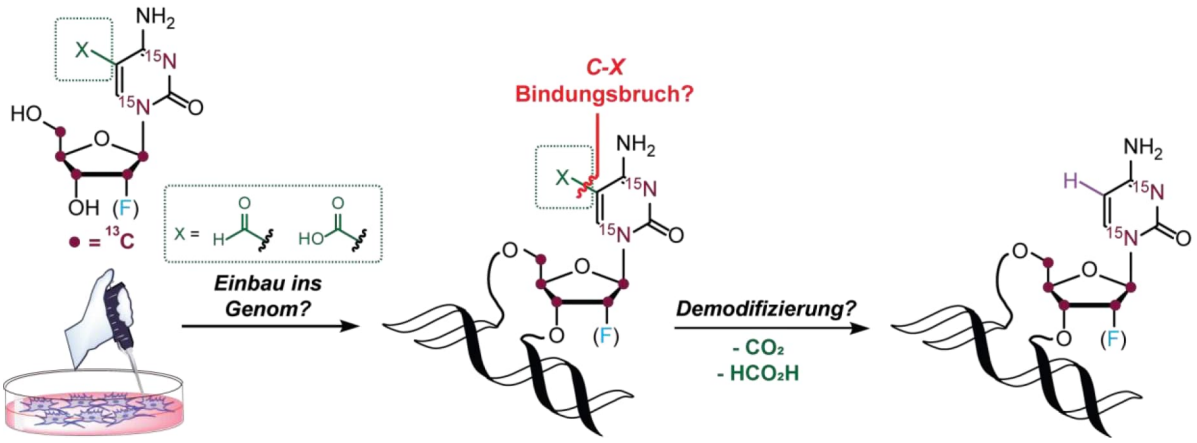


Abbildung 2.1: Schematische Darstellung der Verfütterungsstudien von isotopenmarkierten und/oder 2'-fluorierten Testmolekülen, Derivaten von fdC und cadC, für die Untersuchung der Beteiligung dieser DNA-Modifikationen an der aktiven DNA-Demethylierung durch einen direkten C-C-Bindungsbruch. Die Testmoleküle wurden an Zellkulturen verfütttert und können ins Genom der untersuchten Zellen eingebaut und ensprechend metabolisiert werden. Nach Isolation der genomischen DNA wurde diese auf Nukleosidebene verdaut und per UHPLC-MS/MS analyisert.

Hierzu sollten 2'-(*R*)-fluorierte bzw. isotopenmarkierte Testmoleküle für die zellbiologischen Experimente hergestellt werden. Diese Testmoleküle sollten über einen neuen Syntheseweg zugänglich sein, der das intermediäre Vorkommen von 2'-F-dC sowie des isotopenmarkierten Analogons vermeidet, da diese Verbindungen die Produkte der untersuchten genomischen

Prozesse (Deformylierung und fdC und Decarboxylierung von cadC zu dC) darstellen. Es ist essenziell, das Vorhandensein der dC-Derivate in der Synthesestrategie für die Verfütterungsexperimente zu umgehen. Bereits eine Kontamination im Spurenbereich könnte die Ergebnisse der zellbiologischen Studien verfälschen, da die unmodifizierten dC-Derivate im Vergleich zu den 5-modifizierten dC Derivaten bevorzugt ins Genom eingebaut werden. Für die Untersuchung der genomischen Deformylierung sollte das 2'-fluorierte fdC Derivat sowie ein isotopenmarkiertes Analogon hergestellt werden. Durch den Einsatz der 2'-Fluormarkierung der Ribose wird das Testmolekül um 18 u schwerer als das natürlich vorkommende fdC. Somit lässt sich das verfütterte Testmolekül, sowie deren Metabolismusprodukte mittels massenspektrometrischer Techniken sehr gut analysieren und von dem natürlich vorkommenden fdC unterscheiden. Durch die Bioisosterie des Fluors und Wasserstoffs ergeben sich für die Reaktivität der 2'-fluorierten dC Derivate zwei Vorteile. Auf der einen Seite besitzen die 2'-fluorierten Derivate eine mit den natürlichen dC Modifikationen identische Reaktivität in enzymatischen Reaktionen, die an der Nukleobase ablaufen. Auf der anderen Seite unterliegen die 2'-fluorierten Derivate keiner Basenexzisionsreparatur vermittelten Entfernung aus dem Genom, da das bei der Spaltung der *N*-glykosidischen Bindung während der BER intermediär entstehende Oxocarbenium Ion durch die 2'-Fluorierung stark destabilisiert wäre. Durch den Einsatz von 2'-F-fdC lässt sich somit gezielt und selektiv die direkte C-C-Bindungsbruchreaktivität von fdC untersuchen. Somit stellt dieses Testmolekül ein ideales Modellsystem dar. Neben dem fluorierten Testmolekül sollte ein isotopenmarkiertes fdC Derivat synthetisiert werden, welches mit ^{13}C-Markierungen an der Ribose und ^{15}N-Markierungen an der Nukleobase versehen ist. Das isotopenmarkierte Testmolekül sollte auch in einer neuartigen Synthesestrategie dargestellt werden, welche nicht über das dC Intermediat verläuft, um jegliche potenzielle Kontaminationen bei den Verfütterungsexperimenten zu vermeiden. Die chemisch modifizierten Derivate von fdC sollten im Anschluss in einer Reihe von Verfütterungsexperimenten an unterschiedliche Zelllinien eingesetzt werden. Hierbei sollte untersucht werden, wie stark die Deformylierungsreaktion in unterschiedlichen Systemen ausgeprägt ist, sowie von welchen Faktoren (z. B. Zellteilungsrate) sie abhängig ist. Außerdem sollten Verfütterungsexperimente mit *Knockout*-Zelllinien durchgeführt werden, um potenzielle Proteine zu identifizieren, die an der Deformylierungsreaktion von fdC beteiligt sein könnten. Neben der Beteiligung von fdC an der aktiven DNA-Demethylierung mittels eines direkten C-C-Bindungsbruchs wurde eine analoge Reaktivität für cadC postuliert.[93, 133] Durch den Einsatz von modifizierten cadC-Testmolekülen in zellbiologischen Verfütterungsexperimenten sollte untersucht werden, ob cadC *in vivo* auf genomischer Ebene zu dC decarboxyliert und so möglicherweise auch an der aktiven Demethylierung von mdC beteiligt ist. Dazu sollten zunächst 2'-(*R*)-fluorierte cadC Derivate

entwickelt werden, die eine gute Zellpermeabilität aufweisen und im Anschluss ins Genom eingebaut werden. Durch die unter physiologischen Bedingungen negativ geladene Carboxylatgruppe von cadC, ist die Permeabilität des Nukleosids durch die Zellmembran problematisch. Dieses Problem sollte durch Synthese von cadC Derivaten (*prodrugs*), welche eine Maskierung der Ladung durch Veresterung der Carboxylatgruppe aufweisen, gelöst werden. Die cadC Ester sollten in der Zelle enzymatisch hydrolysierbar sein und im intrazellulären Raum das Testmolekül 2'-F-cadC freisetzen. Falls die Verfütterung der cadC *prodrugs* nicht zu zufriedenstellendem Einbau von 2'-F-cadC führen sollte, sollte ein weiterer Ansatz verfolgt werden, um ein cadC Testmolekül in die Zelle zu transportieren. Durch die Kombination eines 2'-F-cadC Triphosphats mit einem Cyclodextrin-basiertem Nukleotidtransporter sollten zum einen die für den Einbau ins Genom benötigten enzymatischen Phosphorylierungschritte umgangen und zum anderen das Nukleotid direkt in die Zelle transportiert werden. Die Testmoleküle sollten an mESCs verfüttert werden, wo sie in das Genom der Zellen eingebaut werden und im Zuge der aktiven Demethylierung zu 2'-F-dC umgewandelt werden können. Zur exakten Quantifizierung der fluorierten Testmoleküle und deren Metabolismusprodukte im Genom sollten stabile Isotopologe synthetisiert werden, welche als interne Standards für die massenspektrometrische Analyse und Quantifizierung fungieren sollten. Für Untersuchungen des strukturellen und dynamischen Einflusses von cadC auf den DNA-Doppelstrang sollten in Zusammenarbeit mit *Dr. Petra Rovó* NMR Experimente kurzer synthetischer Oligonukleotide, die cadC enthalten, durchgeführt werden.

Um ein tieferes Verständnis für die Reaktivität der oxidierten mdC Modifikationen fdC und cadC zu erhalten, sollte eine Struktkur-Aktivität-Beziehungsstudie durchgeführt werden. Hierfür sollten chemisch modifizierte Analoga von fdC und cadC synthetisiert werden, welche ähnliche stereoelektronische Eigenschaften und eine vergleichbare Reaktivität aufweisen, jedoch eine andere funktionelle Gruppe besitzen, synthetisiert werden. Diese Nukleoside sollten mit funktionellen Gruppen, die analog zum Aldehyd oder der Carbonsäure ein Keton, Nitril, Sulfoxid oder Carbonsäureamid enthalten. Diese chemisch modifizierten Testmoleküle sollten in zellbiologischen Verfütterungsexperimenten eingesetzt werden und deren Einbau und genomische Reaktivität untersucht und mit der Reaktivität von fdC und cadC verglichen werden.

3 VERÖFFENTLICHTE ARBEITEN

3.1 Deformylation of 5-formylcytidine in different cell types

E. Korytiaková*****, E. Kamińska*, M. Müller, T. Carell, *Angew. Chem. Int. Ed*. **2021**, *60*, 16869 – 16873.

*: Diese Autoren haben zu gleichen Teilen beigetragen.

Prolog

Nachdem von *Iwan et al.*[135] gezeigt werden konnte, dass eine direkte Deformylierung von genomischem fdC zu dC im Genom von mausembryonalen Stammzellen abläuft, stellte sich zum einen die Frage, wie weit dieser Prozess in anderen Zellsystemen ausgeprägt ist und zum anderen, unter welchen Bedingungen bzw. Zellproliferationsrate der C-C-Bindungsbruch besonders effizient abläuft. Im Rahmen dieser Studie wurde eine neuartige Synthesestrategie von 2'-F-fdC entwickelt, die das zu verfütternde Testmolekül in extrem hoher Reinheit lieferte. Der Syntheseweg verlief ausschließlich über 2'-F-Desoxyuridin Derivate und stellte somit sicher, dass mögliche Kontaminationen durch 2'-F-dC absolut ausgeschlossen wurden. Im Rahmen dieser Studie wurde 2'-F-fdC in Verfütterungsstudien an eine Vielzahl an Stammzellen sowie somatischen Zelltypen getestet und der Zusammenhang zwischen Deformylierungsaktivität und Zellproliferation untersucht. Es konnte gezeigt werden, dass der Deformylierungsprozess von fdC in allen untersuchten Zellsystemen unterschiedlich stark ausgeprägt ist und invers mit der Zellproliferationsrate korreliert. Außerdem wurde die aktive Demethylierung in Neurogenin induzierbaren iPS Zellen, die zu bipolaren Neuronen differenziert werden, untersucht. In diesem zellulären System hat die Zellteilung einen *steady state* erreicht und Deformylierungsaktivität wurde so in einem statischen System, unabhängig von der Zellteilung untersucht. In dieser Studie konnte gezeigt werden, dass der Großteil an Deformylierung in weniger als 48 h durchläuft und nach dem Zeitpunkt, bei dem die Zellteilung stillgelegt wurde, wurden stabile Level an 2'-F-fdC im Genom erreicht.

Autorenbeitrag

Bei dieser Studie war ich für die Entwicklung und Synthese der fluorierten Nukleosidderivate zuständig, die in zellbiologischen Verfütterungsexperimeten eingesetzt wurden. Außerdem war ich für die Planung der Experimente, Verdau der genomischen DNA und

massenspektrometrische Analyse der Proben, sowie Auswertung der Ergebnisse zuständig. *Ewelina Kamińska* hat zellbiologische Fütterungsexperimente durchgefürt und war an der Analyse der massenspektrometrischen Daten beteiligt. Alle Experimente und erhaltene Daten wurden in enger Zusammenarbeit mit *Markus Müller* und *Thomas Carell* geplant und diskutiert.

Lizenz

Kopie der Open Access Publikation auf Basis der Creative Commons Attribution-NonCommercial License (CC BY-NC 4.0), die den nicht-kommerziellen Abdruck des Artikels erlaubt.

Artikel einsehbar unter:

https://onlinelibrary.wiley.com/doi/full/10.1002/anie.202107089

Lizenz einsehbar unter:

https://creativecommons.org/licenses/by-nc/4.0/

Die *Supporting Information* dieser Veröffentlichung ist im Anhang (Kapitel 8.1) abgebidet.

Check for updates

How to cite: *Angew. Chem. Int. Ed.* **2021**, *60*, 16869–16873
International Edition: doi.org/10.1002/anie.202107089
German Edition: doi.org/10.1002/ange.202107089

Nucleosides

Deformylation of 5-Formylcytidine in Different Cell Types

*Eva Korytiaková[+], Ewelina Kamińska[+], Markus Müller, and Thomas Carell**

Abstract: *Epigenetic programming of cells requires methylation of deoxycytidines (dC) to 5-methyl-dC (mdC) followed by oxidation to 5-hydroxymethyl-dC (hmdC), 5-formyl-dC (fdC), and 5-carboxy-dC (cadC). Subsequent transformation of fdC and cadC back to dC by various pathways establishes a chemical intra-genetic control circle. One of the discussed pathways involves the Tdg-independent deformylation of fdC directly to dC. Here we report the synthesis of a fluorinated fdC feeding probe (F-fdC) to study direct deformylation to F-dC. The synthesis was performed along a novel pathway that circumvents any F-dC as a reaction intermediate to avoid contamination interference. Feeding of F-fdC and observation of F-dC formation in vivo allowed us to gain insights into the Tdg-independent removal process. While deformylation was shown to occur in stem cells, we here provide data that prove deformylation also in different somatic cell types. We also investigated active demethylation in a non-dividing neurogenin-inducible system of iPS cells that differentiate into bipolar neurons.*

Methylation of dC to 5-methyl-dC (mdC) by Dnmt enzymes in higher vertebrates is a means to control the transcriptional activity of genes.[1–4] Reversing the methylation requires oxidation of mdC to hmdC, fdC, and cadC.[5–9] These oxidized bases either block maintenance methylation (hmdC), which leads to global replication-dependent demethylation, or they are removed as nucleosides by the action of the thymine DNA glycosylase (Tdg) repair enzyme (fdC and cadC).[10–16] Alternatively, a direct deformylation and decarboxylation of fdC and cadC to dC may occur.[17–19] The advantage of a direct mechanism, in contrast to the Tdg-repair process, is that it proceeds without the formation of DNA strand breaks.[20,21] These are known to cause genome instability.[22] We could recently show that fdC is deformylated to dC in stem cells.[17] We have also shown that the nucleophilic attack at the C6 position of fdC is necessary to cleave the C–C bond in a deformylation reaction.[18] Today, it is however still unclear to what extent the deformylation reaction occurs in various cell types. It is also unknown whether this process is influenced by replication. In order to investigate these questions, we performed a series of metabolic feeding studies using a 2′-fluorinated fdC derivative obtained by a synthetic pathway that circumvents F-dC as an intermediate. This probe molecule was used to label genomes in different somatic cell types originating from healthy and cancer tissue. For further investigation of active demethylation upon cell differentiation and its dependence on cell replication, additional studies in an iPS cell line system were established.

In order to trace the incorporated fdC probe as well as its deformylation product and to recognize it in front of the background of the abundant genomic dC, a chemical modification of the probe molecule is needed. This can be achieved by incorporating stable isotope labels (^{13}C, ^{15}N, ^{2}H) into the fdC ribose or the fdC heterocycle (not the cleaved-off formyl group). Another elegant way to achieve the labeling of the nucleoside is to tag the ribose with a 2′-F atom, which has the additional advantage that the base of this modified nucleoside cannot be removed by base excision repair.[23,24] In addition, the 2′-fluorinated dC-derivatives are recognized by Dnmt and Tet-enzymes, which means that they can be epigenetically modified (methylated, oxidized, and demethylated). Thus, a feeding probe such as 2′-F-fdC (**1**) is an ideal tool molecule to study active demethylation of mdC. We proved this principle in an in vivo cell system using Tdg knockout mESCs and the respective wildtype line. While the natural genomic fdC levels increase dramatically in the absence of Tdg, the 2′-fluorinated nucleoside values stay the same in wt and KO (Figure SI-6) cells, showing that Tdg is unable to remove 2′-F-fdC as expected. We also investigated the digestion efficiency of DNA strands containing 2′-fluorinated nucleosides and found only a small difference compared to non-fluorinated nucleosides under our conditions.[23,24]

Since 2′-fluorinated nucleosides do not undergo Tdg-mediated repair, feeding of 2′-F-fdC (**1**) to different cell lines gives higher levels of incorporation into the respective genomes (Figure 1 A). Upon deformylation of 2′-F-fdC (**1**), 2′-F-dC (**2**) is formed. This compound can be easily distinguished from genomic dC using an extremely sensitive, non-high resolution UHPLC-MS/MS system (Figure 1 B) because of the different retention times, the difference in molecular weight (+18 amu), and a specific fragmentation pattern.

For feeding experiments, it is essential to synthesize 2′-F-fdC (**1**) via a pathway that avoids 2′-F-dC (**2**) as a synthesis intermediate. Even the smallest contamination of 2′-F-fdC (**1**) with 2′-F-dC (**2**) can lead to false-positive results. 2′-F-dC (**2**) is taken up by cells as well and incorporated into the genome

[*] M. Sc. E. Korytiaková,[+] E. Kamińska,[+] Dr. M. Müller, Prof. Dr. T. Carell
Department of Chemistry
Ludwig-Maximilians-Universität München
Butenandtstrasse 5–13, 81377 Munich (Germany)
E-mail: Thomas.Carell@lmu.de
Homepage: http://www.carellgroup.de

[+] These authors contributed equally to this work.

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/anie.202107089.

© 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

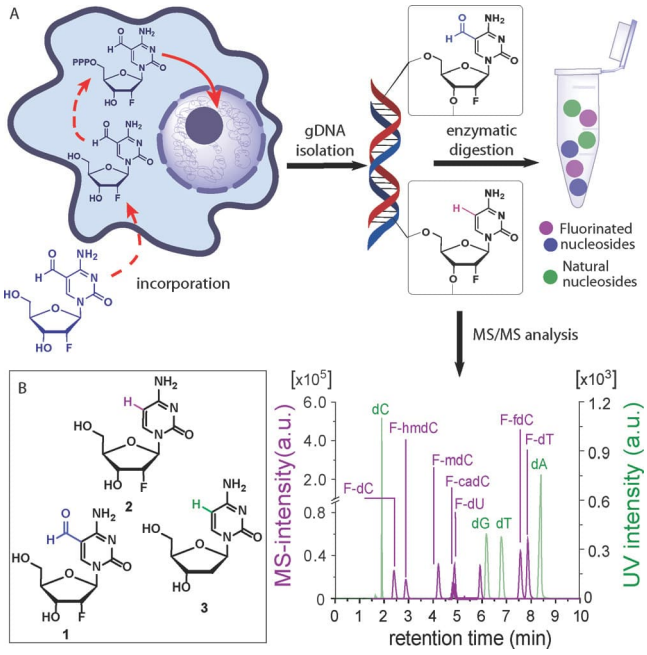
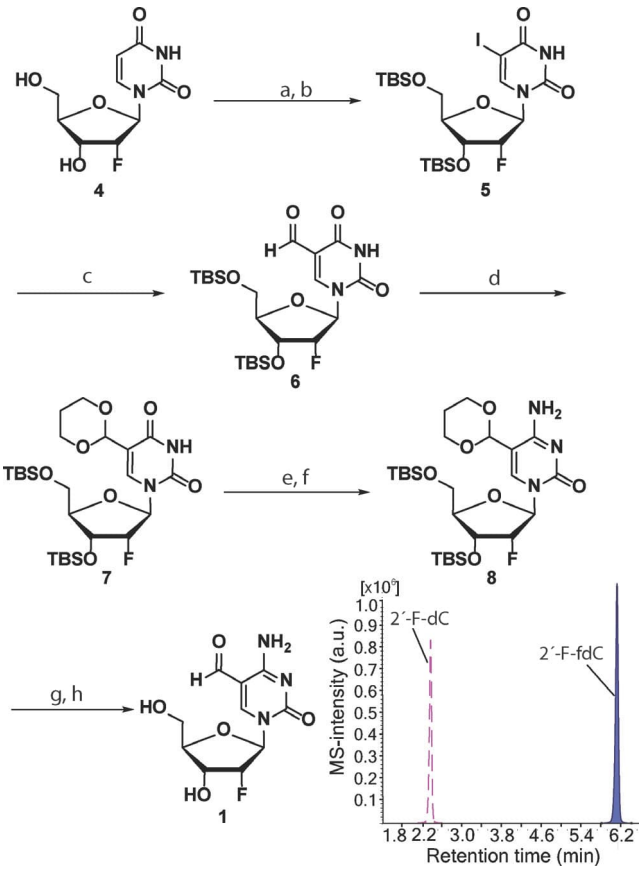


Figure 1. A) Experimental setup and intra-genomic 2′-F-fdC deformylation reaction. The nucleoside is fed to the cells, where it undergoes subsequent phosphorylation in the cytosol, followed by incorporation into gDNA. The extracted gDNA contains the incorporated 2′-F-fdC as well as the possible deformylation product 2′-F-dC. MS/MS analysis was performed after enzymatic digestion of the gDNA. UHPLC-MS/MS chromatogram of all nucleosides present in gDNA. Green peaks show the UV intensity of the compounds, while purple signals represent MS-intensity values. B) The structures of the probe molecules: 2′-F-fdC (**1**), 2′-F-dC (**2**), and natural dC (**3**).



Scheme 1. Synthesis of 2′-F-fdC (**1**) probe molecule and UHPLC experimental data showing the high purity of **1**. a) TBSCl, imidazole, pyridine, rt, 16 h, 91%. b) I_2, $(NH_4)_2[Ce(NO_3)_6]$, MeCN, 60°C, 1 h, 81%. c) $HSnBu_3$, $Pd_2(dba)_3{\cdot}CHCl_3$, PPh_3, 3.5 bar CO, 60°C, 18 h, 85%. d) 1,3-propanediol, $CH(OEt)_3$, *p*-TSA, CH_2Cl_2, rt, 60%. e) NaH, 2,4,6-triisopropylbenzolsulfonylchloride, THF, 0°C→rt, 16 h, 60%. f) NH_4OH, 1,4-dioxane, rt, 16 h, 68%. g) 1 M HCl, H_2O/MeCN, rt, 18 h. h) HF/pyridine, EtOAc, rt, 16 h, reversed-phase HPLC 70%.

far more efficiently compared to 2′-F-fdC (**1**).[17] This leads to a strong enrichment of any 2′-F-dC (**2**) impurity and hence potentially to its accumulation during cell division. Therefore, we developed a novel synthesis pathway towards 2′-F-fdC (**1**) exclusively via 2′-F-dU intermediates thereby excluding undesired cytidine-based contaminations (Scheme 1). The synthesis starts with commercially available 2′-F-dU (**4**), which was first TBS-protected (TBS-Cl, imidazole), followed by regioselective iodination at C5 to **5** with iodine and $(NH_4)_2[Ce(NO_3)_6]$ as the oxidant. A subsequent Pd-catalyzed Stille-type formylation (CO, $HSnBu_3$) furnished the 2′-F-fdU compound **6**. Acetal protection of the formyl group (propanediol, TSA) was followed by amination at C4 to generate the acetal-protected 2′-F-fdC derivative **7**. Full deprotection of **7** with 1 M HCl followed by treatment with HF/pyridine to cleave the TBS protection groups furnished the 2′-F-dC-free target compound **1**. Subsequent HPLC purification provided 2′-F-fdC (**1**) in a highly pure form. In order to substantiate that the prepared 2′-F-fdC (**1**) is 2′-F-dC (**2**)-free, we performed a highly sensitive UHPLC-MS/MS experiment (Scheme 1, inset).

The analysis of the synthetized 2′-F-fdC (**1**) was performed using the highest-sensitive triple quadrupole MS detector. Under conditions previously developed in our group, it is able to detect **2** in the low fmol range (Table SI-1).[17] The analysis of the synthesized 2′-F-fdC (**1**) provided no signal between 2.1 and 2.8 min (retention time of potential 2′-F-dC 2.3 min), which confirmed the high purity of the synthesized compound. Most importantly, this new synthesis route fully excludes any contamination of **1** with 2′-FdC (**2**).

To investigate the deformylation rates in various somatic cell lines and to compare them to mouse embryonic stem cells (mESCs) that we had studied before, respective cultures were grown in the presence of 2′-F-fdC (**1**) at a final concentration of 350 μM (Figure 2). The treatment continued for 3 days in order to allow all cell lines, having different doubling rates, to incorporate the fed compound. The cells were subsequently harvested, DNA was isolated and fully digested using an optimized protocol for the chemically modified nucleosides (see SI).[25] The obtained nucleoside mixture was subsequently analyzed by UHPLC-MS/MS. Quantification of the 2′-F-fdC (**1**) and 2′-F-dC (**2**) levels as well as of the remethylated species 2′-F-mdC (**9**) was performed using the previously published isotope dilution mass spectrometry technique (Figure SI-9).[17,26] First, we noted that the 2′-F-fdC probe molecule was incorporated into all tested cells and detectable at levels between 2.1×10^{-7} and 5.2×10^{-7} molecules per dN (Figure SI-1).

 © 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

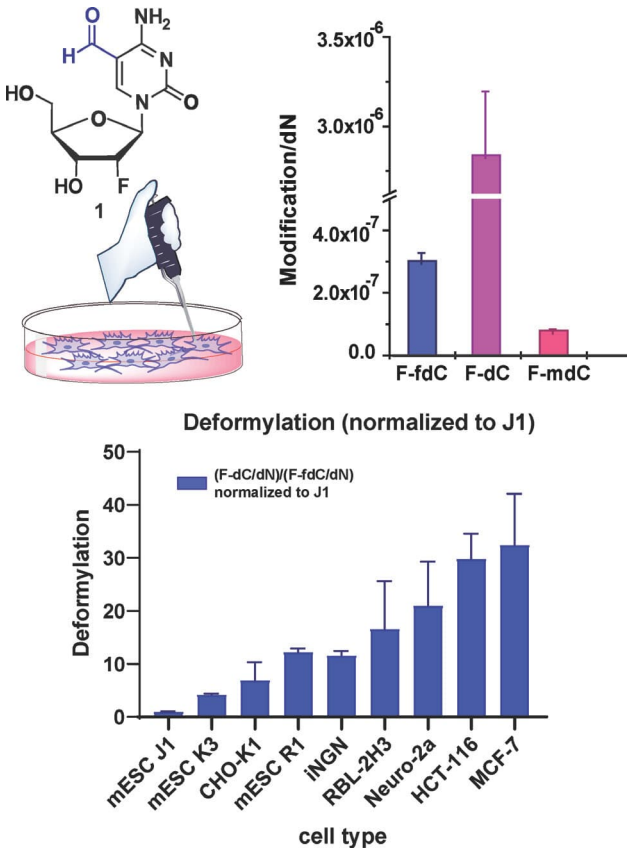


Figure 2. Feeding data of 2′-F-fdC (**1**). The right upper part shows the graph with the amounts of fluorinated nucleosides in gDNA after feeding 2′-F-fdC (**1**) to the J1 cell line. The bottom graph shows the different deformylation efficiencies in different cell lines normalized to the lowest deformylating cell line J1. The error bars represent propagated error based on the standard deviation of all biological triplicates.

As noted earlier, beside the fed compound **1** we also found 2′-F-dC (**2**) and, importantly, 2′-F-mdC (**9**) in all cells (Figure 2).[17] We conclude that the deformylation occurred in all investigated cells and that the newly formed 2′-F-dC (**2**) was subsequently re-methylated by the action of Dnmt enzymes, forming 2′-F-mdC (**9**, Figure 2). This indicates the presence of the chemical genomic cycle dC→mdC→oxdC→dC→mdC in all investigated cell types. Unfortunately, we are not yet able to show the methylation of **2** in the CpG context, as the incorporation of **1** is assumably random. However, the level of methylation of 2′-F-dC (**2**) to 2′-F-mdC (**9**) is with 2–5% within the expected range (Figure SI-8B).[27]

Investigation of the deformylation reaction (turnover of 2′-F-fdC (**1**) to 2′-F-dC (**2**)) in different cell lines showed a major variation of this activity (Figure 2). Most interesting is the observation that the previously studied stem cells J1, despite their high proliferation rate, showed the lowest deformylation activity. In order to analyze all the data, we therefore normalized the 2′-F-fdC (**1**) deformylation level of each cell line to J1 mES cells. When analyzing the rates of deformylation, we noticed that they inversely correlate ($R^2 = 0.83$) with the rates of cell division (Figure SI-2A). In general, the cells that divide the fastest are the mESCs. In contrast, R1, J1, and K3 (between 17 and 19 h doubling time) show the lowest deformylation rates, while the cancer-originating somatic cell lines (with doubling times between 25 and 45 hours) show more efficient deformylation activity. This observation is interesting, as this inverse correlation with cell division could indicate that deformylation is an active process that occurs to a larger extent within resting cells.

We could not correlate the deformylation levels with the activity of Tet enzymes ($R^2 = 0.0005$, based on the levels of genomic natural hmdC, Figure SI-2B). In conclusion, the combined data show that deformylation of 2′-F-fdC (**1**) is abundant in a broad spectrum of cell types. The rate of this process strongly depends on the type of the cell. The observed deformylation process is less efficient in stem cells, while it occurs to high extent in human cancer cell lines that in this study are derived from colon and breast cancer tissue.

We performed a control experiment to evaluate the stability of the fed 2′-F-fdC (**1**) material in the cytosolic nucleoside fraction of the cell and hence under conditions as natural as possible. To this end, the probe **1** was added to the cell culture to achieve its uptake into the cells. Next, after 72 h, we isolated the cytosol containing the soluble nucleosides, extracted the nucleosides and analyzed them by UHPLC-MS/MS to look for the presence of **1** and the deformylated putative product **2** (Figure SI-3A). This experiment allowed us to estimate the stability of **1** and to exclude the possibility that the deformylation reaction occurs in the cytosol. If **2** had formed in the cytosol, it would have been able to incorporate into the DNA, obscuring all discussion of intragenomic deformylation. To our delight, we were unable to detect any **2** in the cytosol, which indicates that **1** is stable and that deformylation does not happen outside the genome. This is true within the limit of detection for 2′-F-dC (**2**), which is, however, as low as 3.0 fmol (Table SI-2). The data together, therefore, again suggest that 2′-F-fdC (**1**) is deformylated after its incorporation into the genome.

Because we detected basic 2′-F-fdC (**1**) levels of 2–5 × 10^{-7}/dN in all investigated cellular systems, we sought to find out if these levels can be reduced to zero when replication ceases. To achieve this aim, we conducted an experiment with iPS cells, which stop replication during differentiation towards neurons. This system allows us to study deformylation independently from replication. For the experiment we used the small molecule-inducible Neurogenin iPS cell line (iNGN cells), which is able to differentiate from the pluripotent state in specific culture conditions (see SI) to cells showing a bipolar neuron-like morphology after only 4 days. Like other neurons, they stop DNA replication and cell division upon maturation, which enables the study of a dynamic process like deformylation without interference from replication-dependent incorporation of fresh material. First, the stability of **1** was again examined analogously to the soluble pool extraction experiment with J1 mESCs. Similar to the previous data we found **1** but no traces of **2** after 72 h in the soluble pool of iNGNs, again showing that **1** is not deformylated under our conditions in the active cell extracts (Figure SI-3B).

To monitor cell division, the cells were fed with a final concentration of 0.5 μM of isotopically labeled dT ($^{13}C_{10}$-$^{15}N_2$-dT, **10**). The incorporation of **10** was then traced over time.

 © 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

After initial decrease of the amount of **10** through cell division, we detected stable levels of **10** in the genome of these cells from 48 h post-induction. This result allows us to conclude that the cells terminate their replication under our conditions after two days from the start of differentiation (Figure SI-4).

We next performed a double feeding experiment with 2′-F-fdC (**1**) and simultaneously with $^{13}C_{10}$-$^{15}N_2$-dT (**10**) as depicted in Figure 3 in order to monitor cell division-dependent wash-out of **10**. This experiment allows us to check how deformylation progresses upon cessation of cell replication. Quantification of genome-integrated modified nucleosides and their metabolites within DNA was performed by UHPLC-MS/MS using the isotope dilution technique (Table SI-3). The deformylation (a rate of change of 2′-F-fdC (**1**) to 2′-F-dC (**2**) relative to dN) was monitored by pulse feeding the cells at 350 μM final concentration of **1** up to 48 h post-induction (start of differentiation) and chasing of the values every 4 hours, from 48 h onwards. The data show that the decrease of 2′-F-fdC (**1**) content within the gDNA (Figure SI-5) continues to a small extent even beyond termination of cell division. We can see that while the levels of **10** in the genome stay constant for the duration of the experiment, the ratios between **2** and **1** increase. Because the total levels of **2** in comparison to **1** are at least 10-fold higher due to the efficiency of turnover, post 48 hours it is not possible to see a clear rise in 2′-F-dC/dN levels. With the knowledge that the deformylation reaction is efficient and occurs already after one hour after incorporation of **1** into the genome, we can assume that the majority of incorporated 2′-F-fdC (**1**) by the 48th hour has already turned over to **2** and **9** (Figure SI-5).[17] Also, the levels of 2′-F-fdC/dN stay constant after 52 h, which further supports the idea that some portion of fdC resides as a (semi)permanent base in the genome. This 2′-F-fdC (**1**) stays hidden in the genome. It seems not accessible for enzymatic processes, potentially because it is situated in condensed chromatin structures.[10,28–33] Following these findings, we decided to perform an in vitro experiment to investigate if histones, with their many amine side chains, could be triggering the deformylation process of fdC.[28] This study, however, was negative, which indicates that more is needed than just the presence of lysine side chains to achieve C–C bond cleavage (Figure SI-7). There is certainly a need for active nucleophiles that are able to attack the C6-position of 2′-F-fdC (**1**) to promote the reaction.[18]

The herein reported data show that the synthesis pathway towards the probe molecule of 2′-F-fdC (**1**) using only 2′-F-dU intermediates ensures the best possible purity of the compound for biological feeding experiments. Feeding of the newly synthesized compound shows that deformylation of 2′-F-fdC (**1**) to 2′-F-dC (**2**) occurs inside the genome and that it is a common process in a variety of cell lines that is inversely correlated with the speed of replication. Within differentiating neurons, we were able to show that the vast majority of deformylation occurs before the 48 hour timepoint, which is the moment when the cells stop replicating. Post-replication, we see stable levels of 2′-F-fdC (**1**) that may be epigenetically relevant, for example, by helping chromatin condensation via Schiff-base formation with the histones.[10,28]

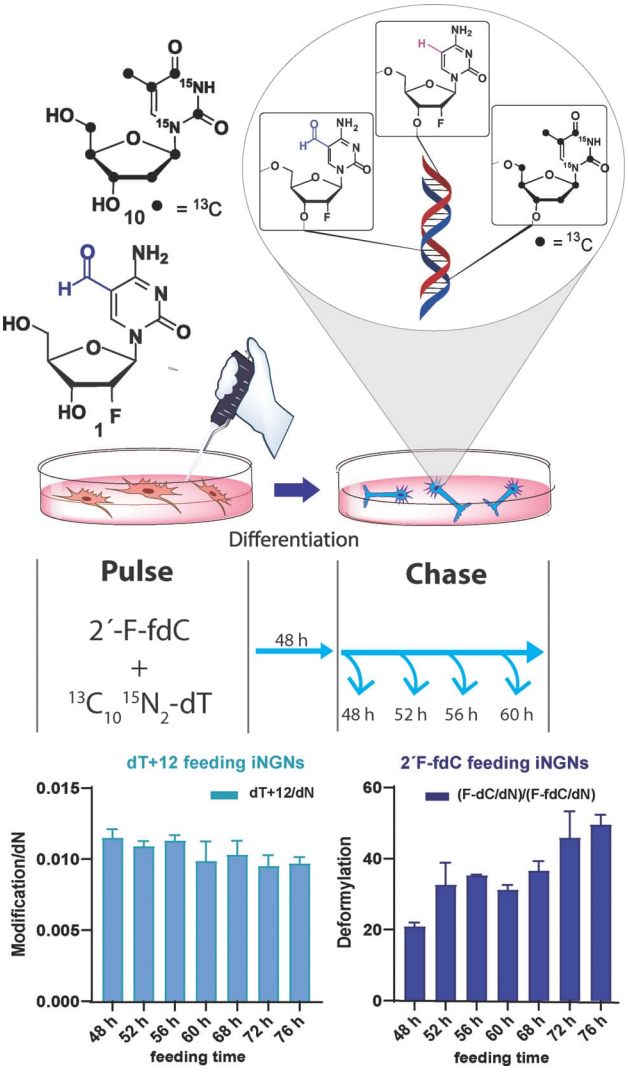


Figure 3. Double feeding experiment of **1** and **10** in iNGN cells, which allow the investigation of the deformylation process depending on cell division. Deformylation data and amounts of incorporated **10** at different time points are shown normalized per dN. The error bars represent propagated error based on standard deviation of each measured technical triplicate.

Acknowledgements

We thank the people who contributed to the success of this project: Dr. Angie Kirchner, Dr. Katharina Iwan, Dr. Leander Runtsch, and Kerstin Kurz. We thank the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for financial support via SFB 1309 (PID 325871075) and SFB 1361 (PID 393547839). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n° EPiR 741912). Additional funding was provided by the Volkswagen Foundation (EvoRib). Open access funding enabled and organized by Projekt DEAL.

 © 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

Conflict of Interest

The authors declare no conflict of interest.

Keywords: active demethylation · C–C bond cleavage · deformylation · epigenetics · formylcytidine

[1] C. C. Chen, K. Y. Wang, C. K. Shen, *J. Biol. Chem.* **2013**, *288*, 9084–9091.
[2] M. Okano, D. W. Bell, D. A. Haber, E. Li, *Cell* **1999**, *99*, 247–257.
[3] E. Li, T. H. Bestor, R. Jaenisch, *Cell* **1992**, *69*, 915–926.
[4] Z. D. Smith, M. M. Chan, T. S. Mikkelsen, H. Gu, A. Gnirke, A. Regev, A. Meissner, *Nature* **2012**, *484*, 339–344.
[5] S. Liu, J. Wang, Y. Su, C. Guerrero, Y. Zeng, D. Mitra, P. J. Brooks, D. E. Fisher, H. Song, Y. Wang, *Nucleic Acids Res.* **2013**, *41*, 6421–6429.
[6] C. B. Mulholland, F. R. Traube, E. Ugur, E. Parsa, E. M. Eckl, M. Schönung, M. Modic, M. D. Bartoschek, P. Stolz, J. Ryan, T. Carell, H. Leonhardt, S. Bultmann, *Sci. Rep.* **2020**, *10*, 12066.
[7] M. Wagner, J. Steinbacher, T. F. Kraus, S. Michalakis, B. Hackner, T. Pfaffeneder, A. Perera, M. Müller, A. Giese, H. A. Kretzschmar, T. Carell, *Angew. Chem. Int. Ed.* **2015**, *54*, 12511–12514; *Angew. Chem.* **2015**, *127*, 12691–12695.
[8] T. Pfaffeneder, B. Hackner, M. Truß, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 7008–7012; *Angew. Chem.* **2011**, *123*, 7146–7150.
[9] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300–1303.
[10] M. Bachman, S. Uribe-Lewis, X. Yang, H. E. Burgess, M. Iurlaro, W. Reik, A. Murrell, S. Balasubramanian, *Nat. Chem. Biol.* **2015**, *11*, 555–557.
[11] V. Valinluck, L. C. Sowers, *Cancer Res.* **2007**, *67*, 946–950.
[12] D.-Q. Shi, I. Ali, J. Tang, W.-C. Yang, *Front. Genet.* **2017**, *8*, 100.
[13] M. Münzel, D. Globisch, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 6460–6468; *Angew. Chem.* **2011**, *123*, 6588–6596.
[14] C. G. Spruijt, F. Gnerlich, A. H. Smits, T. Pfaffeneder, P. W. Jansen, C. Bauer, M. Münzel, M. Wagner, M. Müller, F. Khan, H. C. Eberl, A. Mensinga, A. B. Brinkman, K. Lephikov, U. Müller, J. Walter, R. Boelens, H. van Ingen, H. Leonhardt, T. Carell, M. Vermeulen, *Cell* **2013**, *152*, 1146–1159.
[15] R. Rahimoff, O. Kosmatchev, A. Kirchner, T. Pfaffeneder, F. Spada, V. Brantl, M. Müller, T. Carell, *J. Am. Chem. Soc.* **2017**, *139*, 10359–10364.
[16] T. Fu, L. Liu, Q. L. Yang, Y. Wang, P. Xu, L. Zhang, S. Liu, Q. Dai, Q. Ji, G. L. Xu, C. He, C. Luo, L. Zhang, *Chem. Sci.* **2019**, *10*, 7407–7417.
[17] K. Iwan, R. Rahimoff, A. Kirchner, F. Spada, A. S. Schröder, O. Kosmatchev, S. Ferizaj, J. Steinbacher, E. Parsa, M. Müller, T. Carell, *Nat. Chem. Biol.* **2018**, *14*, 72–78.
[18] A. Schön, E. Kamińska, F. Schelter, E. Ponkkonen, E. Korytiaková, S. Schiffers, T. Carell, *Angew. Chem. Int. Ed.* **2020**, *59*, 5591–5594; *Angew. Chem.* **2020**, *132*, 5639–5643.
[19] S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Höfner, K. T. Wanner, C. Ochsenfeld, T. Carell, *J. Am. Chem. Soc.* **2013**, *135*, 14593–14599.
[20] C. Kunz, F. Focke, Y. Saito, D. Schuermann, T. Lettieri, J. Selfridge, P. Schär, *PLoS Biol.* **2009**, *7*, e1000091.
[21] S. C. Wu, Y. Zhang, *Nat. Rev. Mol. Cell. Biol.* **2010**, *11*, 607–620.
[22] K. W. Caldecott, *Nat. Rev. Genet.* **2008**, *9*, 619–631.
[23] A. S. Schröder, E. Parsa, K. Iwan, M. Wallner, S. Serdjukow, T. Carell, *Chem. Commun.* **2016**, *52*, 14361–14364.
[24] A. S. Schröder, O. Kotljarova, E. Parsa, K. Iwan, N. Raddaoui, T. Carell, *Org. Lett.* **2016**, *18*, 4368–4371.
[25] F. Yuan, Y. Bi, J.-Y. Zhang, Y.-L. Zhou, X.-X. Zhang, C.-X. Song, *RSC Adv.* **2019**, *9*, 29010–29014.
[26] F. R. Traube, S. Schiffers, K. Iwan, S. Kellner, F. Spada, M. Müller, T. Carell, *Nat. Protoc.* **2019**, *14*, 283–312.
[27] A. Meissner, T. S. Mikkelsen, H. Gu, M. Wernig, J. Hanna, A. Sivachenko, X. Zhang, B. E. Bernstein, C. Nusbaum, D. B. Jaffe, A. Gnirke, R. Jaenisch, E. S. Lander, *Nature* **2008**, *454*, 766–770.
[28] E. A. Raiber, G. Portella, S. Martínez Cuesta, R. Hardisty, P. Murat, Z. Li, M. Iurlaro, W. Dean, J. Spindel, D. Beraldi, Z. Liu, M. A. Dawson, W. Reik, S. Balasubramanian, *Nat. Chem.* **2018**, *10*, 1258–1266.
[29] M. Su, A. Kirchner, S. Stazzoni, M. Müller, M. Wagner, A. Schröder, T. Carell, *Angew. Chem. Int. Ed.* **2016**, *55*, 11797–11800; *Angew. Chem.* **2016**, *128*, 11974–11978.
[30] C. X. Song, K. E. Szulwach, Q. Dai, Y. Fu, S. Q. Mao, L. Lin, C. Street, Y. Li, M. Poidevin, H. Wu, J. Gao, P. Liu, L. Li, G. L. Xu, P. Jin, C. He, *Cell* **2013**, *153*, 678–691.
[31] F. Neri, D. Incarnato, A. Krepelova, S. Rapelli, F. Anselmi, C. Parlato, C. Medana, F. Dal Bello, S. Oliviero, *Cell Rep.* **2015**, *10*, 674–683.
[32] E.-A. Raiber, P. Murat, D. Y. Chirgadze, D. Beraldi, B. F. Luisi, S. Balasubramanian, *Nat. Struct. Mol. Biol.* **2015**, *22*, 44–49.
[33] M. W. Kellinger, C. X. Song, J. Chong, X. Y. Lu, C. He, D. Wang, *Nat. Struct. Mol. Biol.* **2012**, *19*, 831–833.

Manuscript received: May 27, 2021
Accepted manuscript online: June 10, 2021
Version of record online: June 24, 2021

 © 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

3.2 Intragenomic decarboxylation of 5-carboxycytidine

E. Kamińska*, E. Korytiaková*, A. Reichl, M. Müller, T. Carell, *Angew. Chem. Int. Ed.* **2021**, *60,* 23207 –23211.

*: Diese Autoren haben zu gleichen Teilen beigetragen.

Prolog

Es wurde in mehreren Studien gezeigt, dass die epigenetisch relevante Modifikation fdC auf Oligonukleotidebene[133] als auch im Genom[135] zu dC deformyliert und dieser Prozess in einer Vielzahl von Stammzellen sowie somatischen Zellen ausgeprägt ist.[177] Die Deformylierung von fdC konnte in diesen Studien sowohl *in vitro* als auch *in vivo* gezeigt werden. Für cadC wurde eine ähnliche genomische Demodifizierung als Teil der aktiven Demethylierung, eine Decarboxylierungssreaktion zu dC, postuliert. Diese Reaktivität wurde bisher nur *in vitro*[93, 133, 136-137] beobachtet, konnte aber nicht auf genomischer Ebene gezeigt werden. Im Rahmen dieser Arbeit wurden 2'-fluorierten dC Derivate in metabolischen Markierungsexperimenten eingesetzt um die Decarboxylierung von cadC im Genom zu untersuchen. Als Testmoleküle wurden 2'-fluorierte Nukleosidtriphosphate eingesetzt, die den Einbau des cadCs ins Genom erleichtern sollten, da so enzymatische Phosphorylierungsschritte für den Einbau nicht notwendig sind. Die 2'-fluorierten Nukleotide wurden zusammen mit einem von *Zawada et al.*[178] entwickelten Cyclodextrin basierten Nukleotidtransporter an Zellkulturen verfüttert. Auf dieser Weise ist es zum ersten Mal gelungen, ein cadC Derivat, 2'-F-cadC ins Genom einzubauen und eine *in vivo* Decarboxylierungsstudie auf genomischer Ebene durchzuführen. Es konnte gezeigt werden, dass cadC im Genom von Stamm- und somatischen Zellen zu dC decarboxyliert und dass dieser direkter C-C-Bindungsbruch zusammen mit der Deformylierung von fdC einen Teil der aktiven DNA-Demethylierung darstellen könnte.

Autorenbeitrag

Bei dieser Studie war ich für die Entwicklung und Synthese der fluorierten Nukleosidderivate für Verfütterungsexperimenten und die Synthese der isotopenmarkierten Nukleoside, die als massenspektrometrische Standards eingesetzt wurden, zuständig. Außerdem war ich für die Entwicklung der massenspektrometrischen Methoden für die genomische DNA, für die Planung der Experimente, Verdau der genomischen DNA und massenspektrometrische Analyse der Proben, sowie Auswertung der Ergebnisse zuständig. *Ewelina Kamińska* hat zellbiologische Fütterungsexperimente durchgefürt und war an der Analyse der

massenspektrometrischen Daten beteiligt. *Andreas Reichl* hat *in vitro* Experimente durchgeführt und war an der Synthese der Nukleosid Derivate beteiligt. Alle Experimente und erhaltene Daten wurden in enger Zusammenarbeit mit *Markus Müller* und *Thomas Carell* geplant und diskutiert.

Lizenz

Kopie der Open Access Publikation auf Basis der Creative Commons Attribution-NonCommercial License (CC BY-NC 4.0), die den nicht-kommerziellen Abdruck des Artikels erlaubt.

Artikel einsehbar unter:

https://onlinelibrary.wiley.com/doi/10.1002/anie.202109995

Lizenz einsehbar unter:

https://creativecommons.org/licenses/by-nc/4.0/

Die *Supporting Information* dieser Veröffentlichung ist im Anhang (Kapitel 8.2) abgebildet.

Check for updates

Nucleosides

How to cite: *Angew. Chem. Int. Ed.* **2021**, *60*, 23207–23211
International Edition: doi.org/10.1002/anie.202109995
German Edition: doi.org/10.1002/ange.202109995

Intragenomic Decarboxylation of 5-Carboxy-2′-deoxycytidine

*Ewelina Kamińska[+], Eva Korytiaková[+], Andreas Reichl, Markus Müller, and Thomas Carell**

Abstract: *Cellular DNA is composed of four canonical nucleosides (dA, dC, dG and T), which form two Watson–Crick base pairs. In addition, 5-methylcytosine (mdC) may be present. The methylation of dC to mdC is known to regulate transcriptional activity. Next to these five nucleosides, the genome, particularly of stem cells, contains three additional dC derivatives, which are formed by stepwise oxidation of the methyl group of mdC with the help of Tet enzymes. These are 5-hydroxymethyl-dC (hmdC), 5-formyl-dC (fdC), and 5-carboxy-dC (cadC). It is believed that fdC and cadC are converted back into dC, which establishes an epigenetic control cycle that starts with methylation of dC to mdC, followed by oxidation and removal of fdC and cadC. While fdC was shown to undergo intragenomic deformylation to give dC directly, a similar decarboxylation of cadC was postulated but not yet observed on the genomic level. By using metabolic labelling, we show here that cadC decarboxylates in several cell types, which confirms that both fdC and cadC are nucleosides that are directly converted back to dC within the genome by C–C bond cleavage.*

5-Formyl-dC (fdC) and 5-carboxy-dC (cadC) are nucleosides that are found in significant amounts in neurons and stem cells.[1] They are formed by oxidation of 5-methyl-dC (mdC) by the action of Tet enzymes via 5-hydroxymethyl-dC (hmdC).[2,3] hmdC is found in these genomes in large quantities. The initial methylation of dC to mdC is performed by the dedicated methyltransferases Dnmt1, -3a and -3b.[4–7] The higher oxidized mdC derivatives, fdC and cadC are known to be removed by the repair glycosylase Tdg, which cleaves the glycosidic bond between the sugar and the corresponding base.[8,9] The result is the formation of abasic sites that are further processed by AP endonuclease, and finally replaced by an unmodified dC (Figure 1 a).[10]

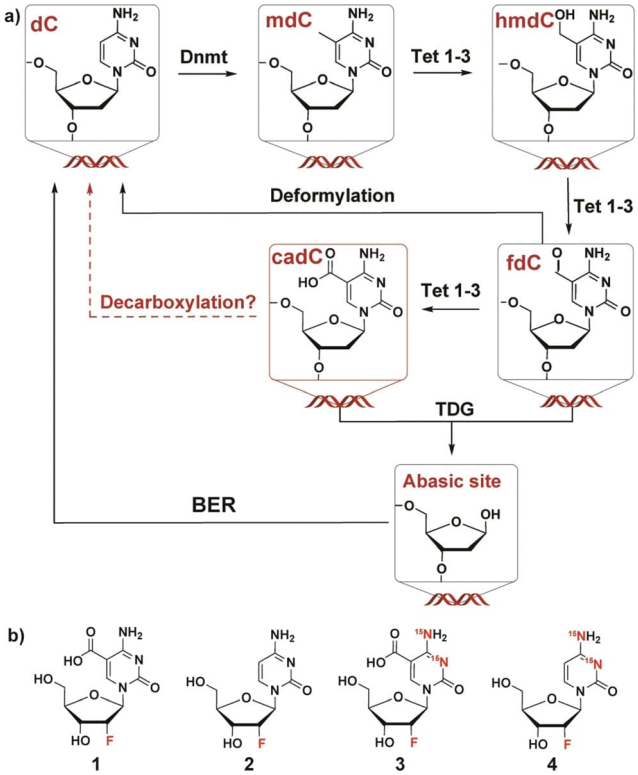


Figure 1. a) Active demethylation pathways via Tdg-mediated excision or direct deformylation and decarboxylation. b) Metabolically fed nucleoside 2′-F-cadC (**1**) and product nucleoside 2-′F-dC (**2**) formed after decarboxylation. MS reference compound [^{15}N]$_2$-2′F-cadC (**3**) and [^{15}N]$_2$-2′F-dC (**4**).[17]

Since the discovery that methylation of dC is followed by oxidation chemistry, it was postulated that fdC and cadC might directly deformylate or decarboxylate to give dC. Chemically, these C–C bond cleavage reactions have the advantage that potentially harmful abasic site intermediates formed during Tdg-mediated active demethylation will not be generated.[11] While for fdC, deformylation was shown to occur in vivo, for cadC just a putative decarboxylation mechanism was postulated in vitro so far.[12,13] It is still unknown whether decarboxylation of cadC occurs in stem cells.[13–16] Here we use our previously described metabolic labelling approach to prove that cadC, if present in the genome of stem and somatic cells, does decarboxylate.[17,18]

Such a direct demethylation reaction in DNA by C–C bond cleavage will give a product that is identical with natural dC. Therefore, it is important to incorporate a reporter nucleoside (cadC*) into the genome of the cells, which generates a decarboxylated dC*. This product must be detectable with high accuracy in the presence of an overwhelming amount of natural dC. In the past we successfully

[*] E. Kamińska,[+] M. Sc. E. Korytiaková,[+] M. Sc. A. Reichl, Dr. M. Müller, Prof. Dr. T. Carell
Department of Chemistry
Ludwig-Maximilians-Universität München
Butenandtstrasse 5–13, 81377 Munich (Germany)
E-mail: Thomas.Carell@lmu.de
Homepage: http://www.carellgroup.de

[+] These authors contributed equally to this work.

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/anie.202109995.

© 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

experimented with 2′-fluoro-labelled nucleosides. Thus, for this study we decided to use 2′F-cadC (**1**) as cadC* (Figure 1b).[17] The synthesis of **1** and its triphosphate was reported by us previously.[19,20] The 2′F-atom ensures that the formed decarboxylated product 2′F-dC (**2**) is readily detectable by UHPLC–MS due to the $m/z = +19$ Da mass shift relative to dC. The specific shift in retention time for fluorinated compounds also allows to distinguish them from canonical dC. This reduces an overlap of 2′F-dC with dC, which avoids ion suppression that makes quantification of even very small quantities of product possible. The 2′F-atom served also a second purpose. Natural cadC is barely detectable in wildtype cells.[8] The levels, however, increase by almost two orders of magnitude, when the base excision repair (BER) pathway is interrupted by knocking out the *TDG* gene (Figure SI-6).[21] The 2′F substitution has the same effect. It blocks the BER process, which leads to higher amounts of detectable, incorporated 2′F-cadC (**1**).[19]

As a reference compound for exact quantification of **1**, we synthesized the isotopologue $[^{15}N]_2$-2′F-cadC (**3**) as an internal standard. Quantification of **2** required the standard $[^{15}N]_2$-2′F-dC (**4**, Figure 1b).

A disadvantage of using 2′F-cadC for studying natural decarboxylation is of course the unnatural character of the nucleoside. In order to investigate if the 2′F-atom influences the decarboxylation reaction, we saturated the C5–C6 bond of the cadC- and 2′F-cadC-methylesters and studied the spontaneous decarboxylation behavior after ester cleavage.[13] To our delight we found that both compounds decarboxylate in a similar manner, which lets us conclude that the effect of the 2′F-atom is small (Figure SI-8). When we started to experiment with 2′F-cadC **1** in cellulo, we learned that it is problematic to label the cellular genome with this compound.[22] The negatively charged 2′F-cadC was only taken up by cells to a small extent and in addition, its intracellular conversion to the triphosphate, as needed for incorporation into the genome, happened to be inefficient as well. It is known that the phosphorylation of cadC by kinases is inefficient.[23]

Although we detected 2′F-cadC (**1**) in the cytosol, we were unable to detect incorporated **1** in the genome. We had to continuously feed the cells for 3 days with **1** to reach detectable, but unquantifiable levels of **1**. After intensive experimentation with different 2′F-cadC delivery methods, we finally succeeded with the help of a cyclodextrin transporter which was modified with a cell-penetrating peptide derivative (Figure 2).[22] This transporter encapsulates nucleoside triphosphates and allows them to be transferred across the cell membrane. Application of this transporter was indeed successful. It allowed us to deliver the **1**-TP with only one 30 min feeding pulse in tricine buffer. After the feeding, the cells were washed, and fresh medium was applied (Figure 2).

In the first experiment, we fed Neuro-2a cells for 30 min with the **1**-TP loaded transporter. The cells were harvested after 24 hours. We then isolated the free nucleotide pool and investigated it regarding the presence of 2′F-cadC (**1**) and 2′F-dC (**2**). In the cytosol, we detected to our delight the presence of delivered **1**. In addition, we saw no **2** (Figure 3a). Together the data show that **1** is a stable compound that, based on our

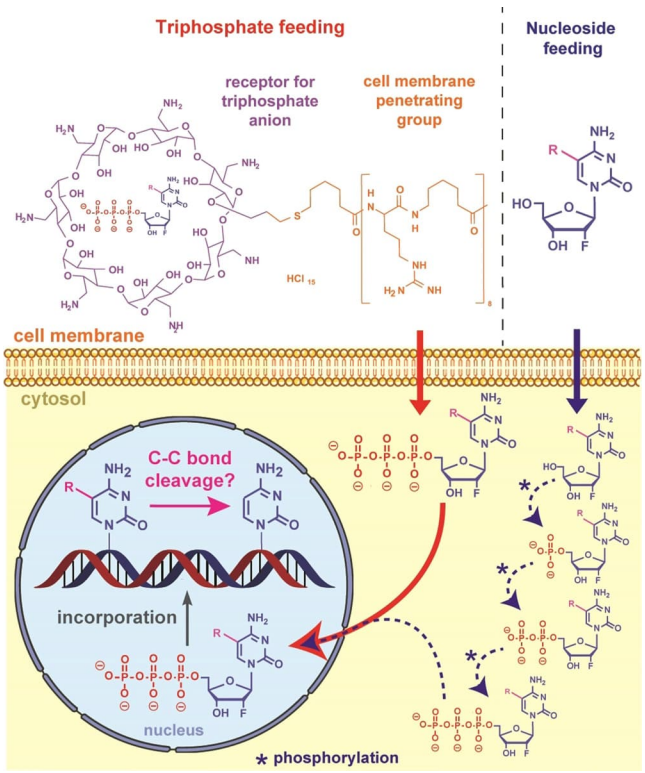


Figure 2. Experimental setup for the inefficient delivery of **1** (blue path) and the accelerated transfer of **1**-TP into cells with the help of the calixarene transporter (red path).

data, does not spontaneously decarboxylate during delivery or under physiological conditions in cells.

Next, we harvested the cells using RLT buffer (Qiagen) supplemented with 400 μM of 2,6-di-*tert*-butyl-4-methylphenol (BHT) and desferoxamine mesylate (DM) as well as β-mercaptoethanol (1:100). The genomic DNA was isolated using a spin column kit (Zymo Research) and digested according to our established method.[18] In brief, the isolated DNA was incubated for 5 min at 95 °C, cooled down on ice and incubated with the Degradase digestion mixture (Zymo Research).

After addition of the isotope-labelled internal standards, the mixture was incubated for 4 h at 37 °C. Next, the samples were diluted with 450 μL of water and were extracted with chloroform. After lyophilization of the aqueous phase, the digested samples were resuspended in water, filtered and analyzed by UHPLC–MS/MS.[18] In order to prove that this digestion method is efficient for 2′F-cadC-containing DNA, we digested a short ssDNA containing a synthetically embedded **1** (Figure SI-1). Best results were obtained with the Degradase mix (Figure SI-1).[24] We next tested the digestion using a 147-base long dsDNA (Widom 601)[25] containing **1**. Using LC–MS, we detected all canonical DNA bases (dA, dC, dG, T) and in addition, **1** at the expected level, confirming that efficient digestion is possible using the Degradase method. In contrast to this, however, we noticed that when we added the same amount of a 2′F-cadC-containing DNA strand to normal genomic DNA, we obtained an astonishingly small signal for 2′F-cadC (**1**). Indeed, only 10% of the expected signal was

 © 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

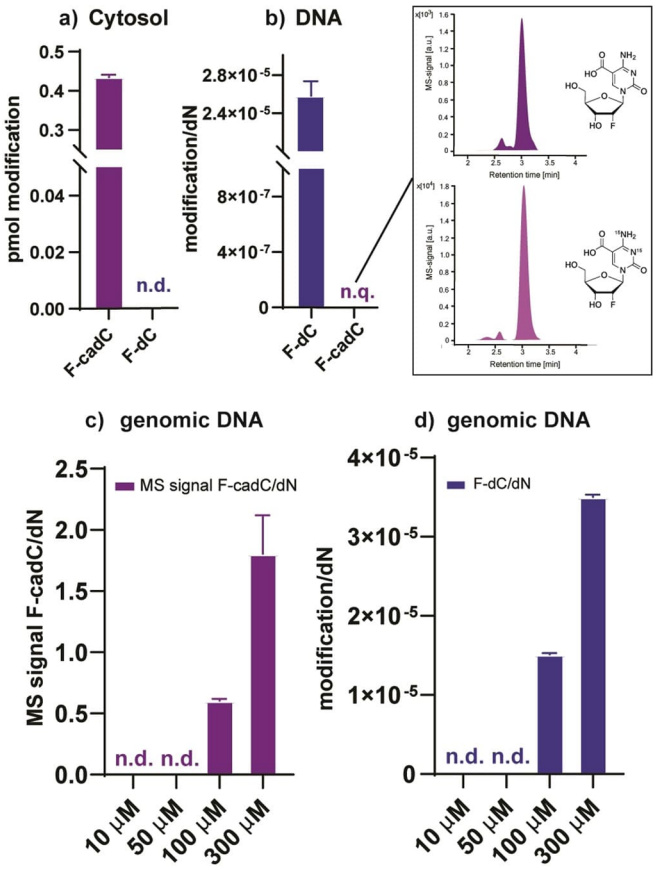


Figure 3. Metabolic feeding of 2′F-cadCTP to Neuro-2a cells. Investigation of the decarboxylation reaction a) in the cytosol and b) in genomic DNA. Dose–response data showing increased MS signal of levels of c) **1** and d) **2** after 24 h with increasing feeding concentration.

detected, arguing that the detectability of **1** is strongly reduced in a complex environment (Figure SI-2). Due to this strong signal suppression, we therefore abstained from exact quantification of **1,** using internal standard **3** in all further experiments.

To investigate the decarboxylation process, we isolated and digested genomic DNA from Neuro-2a cells fed with **1**-TP. In the above-described method, indeed we clearly detected a signal for **1** in the genome of cells fed with **1**-TP based on its retention time, which was identical with the reference compound **3** and its fragmentation pattern, which was indistinguishable from **3** (Figure 3b). To our delight, we also detected the decarboxylated product **2**, which was not seen in the cytosolic fraction. Compound **2** was detected at levels of 2.4×10^{-5} per dN, 24 h post feeding. The result shows that while the decarboxylated product **2** does not form in the cytosol by spontaneous decarboxylation, it is present in the genome, arguing that decarboxylation takes place when 2′F-cadC (**1**) is incorporated into genomic DNA.

Next, we investigated if the decarboxylation signals form in a dose-dependent manner (Figure 3c). Feeding Neuro-2a cells with final concentrations of 100 µM and 300 µM of the **1**-TP-loaded transporter gives indeed an increasing signal for 2′F-dC (**2**, Figure 3d). We also see an increase in the MS signal intensity of 2′F-cadC (**1,** Figure 3c), as expected.

Then, we performed a time-course experiment to investigate at which timepoint decarboxylation becomes detectable (Figure SI-5). We discovered that **2** is detectable only after about 8 h at the earliest, which shows that the C–C bond cleavage has a late onset (Figure SI-5a). As a control experiment, to further exclude that we detect the incorporation of an impurity, we co-fed a 100 µM solution of **1**-TP with a 1% impurity of 2′F-dCTP (**2**-TP, Figure SI-5b). In this experiment, we see again genome-incorporated **1** and **2**, but now, **2** appeared already after 30 min.[17] Together, the data show that the detected 2′F-dC (**2**) is formed from 2′F-cadC (**1**) by C–C bond cleavage.

In the next experiment, we analyzed deformylation of 2′F-fdC (**5**) and decarboxylation of 2′F-cadC (**1**) side by side (Figure 4a). We fed Neuro-2a cells with 2′F-fdC-TP in one experiment and with 2′F-cadC-TP in the second. We used the same amount (400 µM) of both materials and of the transporter. We then exchanged the medium and allowed the cells to recover for 24 h. Afterwards, the cells were harvested, the genomic DNA was isolated, digested and analyzed by UHPLC–MS/MS using the synthetic internal standards **4** as well as [^{15}N]$_2$-2′F-fdC (**6**) for quantification (Figure 4c).

We detected clearly the fed starting material 2′F-fdC (**5**) and again only just traces (due to ion suppression and/or low incorporation) of 2′F-cadC (**1**). Despite this, the amount of the C–C bond cleavage product 2′F-dC (**2**) was higher when we fed **1**-TP compared to **5**-TP (Figure 4a). The exact quantification of **2** in both experiments, using the isotopically labelled material **4**, clearly showed more **2** derived from decarboxylation than deformylation. This interesting result

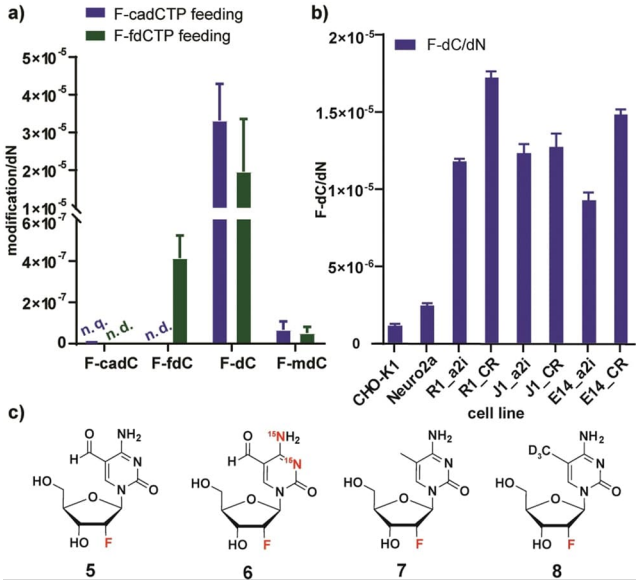


Figure 4. a) Comparison of deformylation and decarboxylation by feeding of 1-triphosphate and 5-triphosphate as well as investigation of the remethylation of the product **2**. b) Decarboxylation in different cells (J1, R1, E14). c) 2′F-fdC (**5**), 2′F-mdC (**7**), as well as the reference compounds [^{15}N$_2$]-2′F-fdC (**6**) and [D$_3$]-2′F-mdC (**8**) needed for exact quantification.[17]

 © 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

shows that we either dramatically underestimate the genomic incorporation of **1** due to the large ion suppression, or that **1** decarboxylates to **2** more efficiently than **5** deformylates. Based on earlier chemical studies and our current knowledge of both processes, we speculate that faster decarboxylation is the factor contributing more.[26]

We next asked the question if the product nucleoside 2′F-dC (**2**) is remethylated. To this end, we quantified the levels of 2′F-mdC (**7**) in both experiments using the internal standard [D_3] 2′F-mdC (**8**). To our delight, we clearly detected **7** in both experiments (Figure 4a). Again, higher amounts were detected from 2′F-cadCTP feeding. This result confirms that after deformylation or decarboxylation, the formed product **2** is remethylated, which suggests the presence of a putative regulatory chemical cycle that starts with methylation of dC by Dnmt enzymes to mdC, followed by oxidation of mdC to hmdC, fdC, and cadC and then deformylation and decarboxylation of fdC and cadC to dC, which could start a new circle.[27] Interestingly, when we measured the levels of **7** formed by methylation of **2** 24 h post **1**-TP feeding, we found only 0.2% methylation. When the same measurement was performed after 72 h, we measured 2.1%, very close to the natural methylation levels of Neuro-2a cells of about 2.5% (Figure SI-7). This result shows that while an early harvest allows us to detect higher amounts of 2′F-dC, the levels of 2′F-mdC are underestimated, potentially because the methyltransferases do not have sufficient time to achieve remethylation.

Finally, we studied how much decarboxylation is taking place in different cells (Figure 4b). We investigated the process in somatic (Neuro-2a, CHO-K1) and in mouse embryonic stem cells (J1, E14, R1). The stem cells were investigated at the pre-implantation and post-implantation stages, obtained by culturing the cells in either a2i or CR media as described in the SI. The data depicted in Figure 4b show that upon feeding **1**-TP at 100 μM over 24 hours, **2** is clearly detectable in all cases but the levels vary between the different cell types. The highest levels of **2** were observed in stem cells cultured under CR conditions. These are the cells representing post-implantation embryos that naturally have the highest methylation levels due to epigenetic reprogramming during cell lineage differentiation.[28] In general, stem cells show an about ten times higher decarboxylation activity than somatic cells, which underpins the potential epigenetic importance of the process.

In summary, the presented data show that next to deformylation of fdC, we also need to consider decarboxylation of cadC as a mechanism for active demethylation. All further efforts now need to be concentrated at finding the cellular entities or circumstances that enable these C–C bond cleavage reactions. Although the here reported data clearly point to the existence of decarboxylation we need to emphasize that cells are complicated entities and we feed an unnatural compound. Without clear identification of the biological entity responsible for the process, we cannot completely rule out that unknown processes other than intragenomic decarboxylation are responsible for the measured data. During the review process of this manuscript Feng and co-workers showed an incorporation of the F-carboxycytosine as a nucleoside and interestingly managed to detect the decarboxylation of cadC to dC much earlier on.[29]

Acknowledgements

We thank the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for financial support via SFB 1309 (PID 325871075) and SFB 1361 (PID 393547839). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n° EPiR 741912). Additional funding was provided by the Volkswagen Foundation (EvoRib) and the DFG priority program SPP 1784 (PID 277203618). Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: active demethylation · carboxycytidine · C–C bond cleavage · decarboxylation · epigenetics

[1] T. Carell, C. Brandmayr, A. Hienzsch, M. Müller, D. Pearson, V. Reiter, I. Thoma, P. Thumbs, M. Wagner, *Angew. Chem. Int. Ed.* **2012**, *51*, 7110–7131; *Angew. Chem.* **2012**, *124*, 7220–7242.
[2] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929–930.
[3] M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930–935.
[4] Z. Li, H. Dai, S. N. Martos, B. Xu, Y. Gao, T. Li, G. Zhu, D. E. Schones, Z. Wang, *Genome Biol.* **2015**, *16,* https://doi.org/10.1186/s13059-015-0685-2.
[5] C.-C. Chen, K.-Y. Wang, C.-K. J. Shen, *J. Biol. Chem.* **2013**, *288*, 9084–9091.
[6] C. Luo, P. Hajkova, J. R. Ecker, *Science* **2018**, *361*, 1336–1340.
[7] M. Okano, D. W. Bell, D. A. Haber, E. Li, *Cell* **1999**, *99*, 247–257.
[8] Y. F. He, B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He, G. L. Xu, *Science* **2011**, *333*, 1303–1307.
[9] L. S. Pidugu, Q. Dai, S. S. Malik, E. Pozharski, A. C. Drohat, *J. Am. Chem. Soc.* **2019**, *141*, 18851–18861.
[10] H. E. Krokan, M. Bjoras, *Cold Spring Harbor Perspect. Biol.* **2013**, *5*, a012583.
[11] M. Wossidlo, V. Sebastiano, K. Lepikhov, M. Boiani, R. Reinhardt, H. Schöler, J. Walter, *EMBO J.* **2010**, *29*, 1877–1888.
[12] S. Schiesser, B. Hackner, T. Pfaffeneder, M. Müller, C. Hagemeier, M. Truss, T. Carell, *Angew. Chem. Int. Ed.* **2012**, *51*, 6516–6520; *Angew. Chem.* **2012**, *124*, 6622–6626.
[13] S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Höfner, K. T. Wanner, C. Ochsenfeld, T. Carell, *J. Am. Chem. Soc.* **2013**, *135*, 14593–14599.
[14] G. Song, G. Wang, X. Luo, Y. Cheng, Q. Song, J. Wan, C. Moore, H. Song, P. Jin, J. Qian, H. Zhu, *Nat. Commun.* **2021**, *12*, 795.
[15] Z. Liutkeviciute, E. Kriukiene, J. Licyte, M. Rudyte, G. Urbanaviciute, S. Klimasauskas, *J. Am. Chem. Soc.* **2014**, *136*, 5884–5887.

 © 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

[16] Y. Feng, N.-B. Xie, W.-B. Tao, J.-H. Ding, X.-J. You, C.-J. Ma, X. Zhang, C. Yi, X. Zhou, B.-F. Yuan, Y.-Q. Feng, *CCS Chem.* **2021**, *3*, 994 – 1008.
[17] K. Iwan, R. Rahimoff, A. Kirchner, F. Spada, A. S. Schröder, O. Kosmatchev, S. Ferizaj, J. Steinbacher, E. Parsa, M. Müller, T. Carell, *Nat. Chem. Biol.* **2018**, *14*, 72 – 78.
[18] F. R. Traube, S. Schiffers, K. Iwan, S. Kellner, F. Spada, M. Müller, T. Carell, *Nat. Protoc.* **2019**, *14*, 283 – 312.
[19] A. S. Schröder, O. Kotljarova, E. Parsa, K. Iwan, N. Raddaoui, T. Carell, *Org. Lett.* **2016**, *18*, 4368 – 4371.
[20] A. S. Schröder, E. Parsa, K. Iwan, M. Wallner, S. Serdjukow, T. Carell, *Chem. Commun.* **2016**, *52*, 14361 – 14364.
[21] R. Rahimoff, O. Kosmatchev, A. Kirchner, T. Pfaffeneder, F. Spada, V. Brantl, M. Müller, T. Carell, *J. Am. Chem. Soc.* **2017**, *139*, 10359 – 10364.
[22] Z. Zawada, A. Tatar, P. Mocilac, M. Buděšínský, T. Kraus, *Angew. Chem. Int. Ed.* **2018**, *57*, 9891 – 9895; *Angew. Chem.* **2018**, *130*, 10039 – 10043.
[23] M. Zauri, G. Berridge, M.-L. Thézénas, K. M. Pugh, R. Goldin, B. M. Kessler, S. Kriaucionis, *Nature* **2015**, *524*, 114 – 118.
[24] F. Yuan, Y. Bi, J.-Y. Zhang, Y.-L. Zhou, X.-X. Zhang, C.-X. Song, *RSC Adv.* **2019**, *9*, 29010 – 29014.
[25] W. J. Lowary, *J. Mol. Biol.* **1998**, *276*, 19 – 42.
[26] E. Korytiaková, E. Kamińska, M. Müller, T. Carell, *Angew. Chem. Int. Ed.* **2021**, *60*, 16869 – 16873; *Angew. Chem.* **2021**, *133*, 17005 – 17010.
[27] A. Parry, S. Rulands, W. Reik, *Nat. Rev. Genet.* **2021**, *22*, 59 – 66.
[28] S. Takahashi, S. Kobayashi, I. Hiratani, *Cell. Mol. Life Sci.* **2018**, *75*, 1191 – 1203.
[29] Y. Feng, J.-J. Chen, N.-B. Xie, J.-H. Ding, X.-J. You, W.-B. Tao, X. Zhang, C. Yi, X. Zhou, B.-F. Yuan, Y.-Q. Feng, *Chem. Sci.* **2021**, *12*, 11322 – 11329.

Manuscript received: July 26, 2021
Accepted manuscript online: August 25, 2021
Version of record online: September 24, 2021

 © 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

3.3 Analysis of an active deformylation mechanism of 5-formyl-deoxycytidine (fdC) in stem cells

A. Schön*, E. Kamińska*, F. Schelter*, E. Ponkkonen, E. Korytiaková, S. Schiffers, T. Carell, *Angew. Chem. Int. Ed.* **2020** *59,* 5591-5594.

*: Diese Autoren haben zu gleichen Teilen beigetragen.

Prolog

Durch die Entwicklung und Einsatz isotopenmarkierter und 2'-fluorierter Desoxycytidin Derivate in zellbiologischen Verfütterungsexperimenten konnte von *Iwan et al*.[135] gezeigt werden, dass es im Genom von Stammzellen zu einer Deformylierung von fdC zu dC kommt. Es wurde auch gezeigt, dass dieser Prozess in einer vielzahl an Zelltypen vorhanden ist.[177] Für den mechanistischen Ablauf der Deformylierungsreaktion kann aus *in vitro* Studien angenommen werden, dass eine derartige Demodifizierungsreaktion durch die Aktivität eines Helfernukleophils abläuft. Die Studie von *Schiesser et al.*[93] zeigte, dass es in Anwesenheit eines Schwefel Nukleophils wie Cystein und Aminosäuren mit basischen Seitenketten wie Arginin oder Histidin zur Decarboxylierung von cadC zu dC auf Oligonukleotidebene kommt. Für diese Reaktion wurde analog zur Wirkungsweise der DNA-Methyltransferasen, ein nukleophiler Angriff des Schwefel Nukleophils an der C-6-Position der Cytosinbase postuliert. Für die Deformylierungsreaktion wird analog die Beteiligung eines Helfernukleophils angenommen. Im Rahmen dieser Studie wurde 6-Aza-fdC, ein fdC Analogon, synthetisiert, bei dem es zu keinem nukleophilen Angriff an die C-6-Position der Nukleobase kommen kann. 6-Aza-fdC wurde zusammen mit 2'-F-fdC an Zellkulturen verfüttert und die Reaktivität dieser Nukleoside im Genom untersucht und verglichen. Die Parallelverfütterung hat gezeigt, dass 2'-F-fdC *in vivo* zum großen Teil zu 2'-F-dC deformyliert. Bei 6-Aza-fdC konnte keine Deformylierungsaktivität zu 6-Aza-dC beobachtet werden. Diese Studie zeigt *in vivo*, dass ein nukleophiler Angriff an die 6-Position von fdC für einen C-C-Bindungsbruch in Form einer Deformylierung notwendig ist.

Autorenbeitrag

In dieser Arbeit war ich für die Entwicklung und Durchführung der Synthese des 2'-F-fdC-Derivats zuständig, welches für *in vivo* Fütterungsexperimente verwendet wurde. *Alexander Schön* entwickelte die Synthesen von 6-Aza-dC-Derivaten und deren Derivatisierungs-produkten. Die zellbiologischen Verfütterungsexperimente wurden von *Ewelina Kamińska*

durchgeführt. Entwicklung der massenspektrometrischen Methoden für die Analyse der genomischen DNA wurde von *Florian Schelter* und *Sarah Schiffers* durchgeführt. *Eveliina Ponkkonen* führte Reaktivitätsstudien mit Oligonukleotiden durch.

Lizenz

Kopie der Open Access Publikation auf Basis der Creative Commons Attribution-NonCommercial License (CC BY-NC 4.0), die den nicht-kommerziellen Abdruck des Artikels erlaubt.

Artikel einsehbar unter:

https://onlinelibrary.wiley.com/doi/full/10.1002/anie.202000414

Lizenz einsehbar unter:

https://creativecommons.org/licenses/by-nc/4.0/

Die *Supporting Information* dieser Veröffentlichung ist einsehbar unter:

https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Fanie.202000414&file=anie202000414-sup-0001-misc_information.pdf

Check for updates

Epigenetics

International Edition: DOI: 10.1002/anie.202000414
German Edition: DOI: 10.1002/ange.202000414

Analysis of an Active Deformylation Mechanism of 5-Formyl-deoxycytidine (fdC) in Stem Cells

*Alexander Schön[+], Ewelina Kaminska[+], Florian Schelter[+], Eveliina Ponkkonen, Eva Korytiaková, Sarah Schiffers, and Thomas Carell**

Dedicated to Dr. Klaus Römer on the occasion of his 80th birthday

Abstract: *The removal of 5-methyl-deoxycytidine (mdC) from promoter elements is associated with reactivation of the silenced corresponding genes. It takes place through an active demethylation process involving the oxidation of mdC to 5-hydroxymethyl-deoxycytidine (hmdC) and further on to 5-formyl-deoxycytidine (fdC) and 5-carboxy-deoxycytidine (cadC) with the help of α-ketoglutarate-dependent Tet oxygenases. The next step can occur through the action of a glycosylase (TDG), which cleaves fdC out of the genome for replacement by dC. A second pathway is proposed to involve C–C bond cleavage that converts fdC directly into dC. A 6-aza-5-formyl-deoxycytidine (a-fdC) probe molecule was synthesized and fed to various somatic cell lines and induced mouse embryonic stem cells, together with a 2′-fluorinated fdC analogue (F-fdC). While deformylation of F-fdC was clearly observed in vivo, it did not occur with a-fdC, thus suggesting that the C–C bond-cleaving deformylation is initiated by nucleophilic activation.*

The nucleobase modification 5-formyl-deoxycytidine (fdC, **1**) is found in stem cells during early development and in the brain.[1–5] These tissues are particularly rich in 5-hydroxymethyl-deoxycytidine (hmdC) from which fdC (**1**) is produced.[6,7] The formation of hmdC and fdC requires oxidation reactions that are performed by α-ketoglutarate-dependent Tet enzymes, with 5-methyl-deoxycytidine (mdC) being the initial starting molecule.[8–10] This cascade of oxidation reactions is a part of an active demethylation process, in which mdC as a silencer of transcription is replaced by unmodified dC.[11] The central molecule that is removed seems to be fdC.[12,13] It can be cleaved out of the genome by a dedicated DNA glycosylase, which creates an abasic site that is further processed, leading to the insertion of an unmodified dC.[14] Because abasic sites are harmful DNA-repair intermediates that can cause genome instability, it was suggested early on that fdC might be directly deformylated to dC by C–C bond cleavage.[15,16] Evidence for the existence of such a direct deformylation process was recently reported.[17] Model studies showed that direct deformylation of fdC and potentially also decarboxylation of 5-carboxy-deoxycytidine (cadC) are indeed possible.[15] Nevertheless, it requires activation of the nucleobase by a nucleophilic addition to the C6 position. For fdC, an additional hydrate formation on the formyl group seems to be necessary, as depicted in Figure 1 A. Although activation with a helper nucleophile is well known as the central mechanistic process during the methylation of dC to mdC by DNA methyltransferases (Dnmts),[18,19] it remains to be confirmed whether such activation occurs in vivo as well.

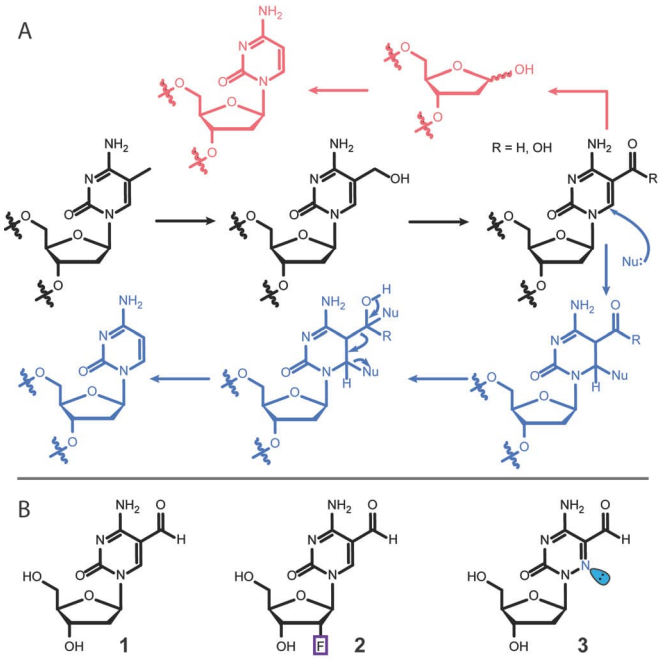


Figure 1. A) The mdC removal pathways that involve oxidation to hmdC, fdC, and cadC followed by either base-excision repair (magenta) or C–C bond cleavage (blue). B) Structures of fdC (**1**) and the two probe molecules **2** and **3** used for this study.

[*] M. Sc. A. Schön,[+] E. Kaminska,[+] M. Sc. F. Schelter,[+] M. Sc. E. Ponkkonen, M. Sc. E. Korytiaková, Dr. S. Schiffers, Prof. Dr. T. Carell
Department of Chemistry, Ludwig-Maximilians Universität München
Butenandtstr. 5–13, 81377 München (Germany)
E-mail: thomas.carell@lmu.de
Homepage: https://www.carellgroup.de

[+] These authors contributed equally to this work.

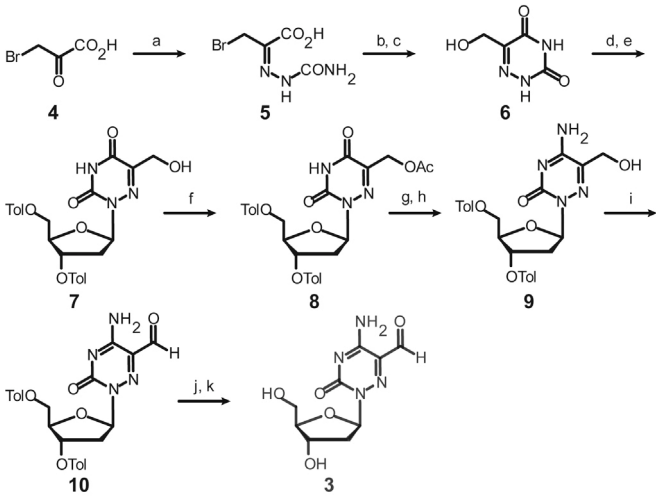
Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/anie.202000414.

© 2020 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited, and is not used for commercial purposes.

In this work, we investigated this hypothesis with two probe molecules, 2′-fluorinated-fdC (F-fdC, **2**) and 6-aza-fdC (a-fdC, **3**). The two compounds were simultaneously fed to different cell types, including primed stem cells. This led to random incorporation of these bases at the "C" sites in the respective genomes. Furthermore, it led to the presence of F-fdC and a-fdC not only at CpG sites. Ultrasensitive UHPLC-QQQ-MS^2 was subsequently used to interrogate the chemical processes that occur at F-fdC and a-fdC in the genomes. The data show that while F-fdC is efficiently deformylated, this does not occur for a-fdC. The only difference between the two nucleobases is the presence of an in-ring nitrogen atom (6-aza atom), which features a lone pair that prohibits nucleophilic addition. These results thus provide strong evidence that nucleophilic activation is the central governing mechanistic event that is required for C–C bond cleavage in vivo.

The fluorinated nucleoside F-fdC (**2**) was recently introduced by us as a deformylation probe.[17] Compound **2** is an antimetabolite that is effectively incorporated into the genomes of growing cells. The 2′-fluoro group is required to block all types of glycosylases, so that base-excision repair is efficiently inhibited. This ensures high levels of F-fdC (**2**) in the genome, as required to observe potential deformylation processes.

The synthesis of the novel nucleoside a-fdC (**3**) is depicted in Scheme 1. The synthesis was started with bromo pyruvic acid (**4**), which we first converted into the semicarbazone **5**, followed by conversion into the acid chloride, subsequent cyclization, and hydrolysis to give hydroxymethylated 6-aza-uracil (**6**).[20] Vorbrüggen nucleosidation with Hoffers' chlorosugar subsequently provided the nucleoside **7** as a mixture of the α- and β-anomers, which could be separated by recrystallization. Next, we acetyl-protected the hydroxymethyl group to give **8**, and then used a standard procedure to convert the U base **8** into the C-derived base **9** by amination of the 4-triazole intermediate with ammonium hydroxide. This led to the concomitant cleavage of the acetyl protecting group. Dess–Martin oxidation of **9** to **10** and final removal of the toluoyl groups furnished the 6-aza-5-formyl-deoxycytidine nucleoside (a-fdC) **3** in a good total yield of 22% with respect to **6** (Supporting Information).



Scheme 1. Synthesis of the probe molecule a-fdC (**3**). a) semicarbazide·HCl, NaOAc, HOAc, H_2O, 0°C to r.t., 2.5 h, 49%. b) pyridine, $SOCl_2$, 80°C, 75 min. c) H_2O, 110°C, 17 h, 74% over 2 steps. d) TMSCl, HMDS, 135°C, 75 min, *then* e) Hoffer's chlorosugar, $CHCl_3$, r.t., 17 h, 56% over 2 steps. f) Ac_2O, pyridine, r.t., 22 h, 96%. g) 1,2,4-triazole, $POCl_3$, NEt_3, MeCN, 0°C to r.t., 18 h, then h) NH_4OH, 1,4-dioxane, 40°C, 5 h, 84%. i) Dess–Martin periodinane, CH_2Cl_2, −15°C to r.t., 1 h, 89%. j) NaOMe, MeOH, benzene, r.t., 1.5 h, then k) reversed-phase HPLC, 54%.

Compound **3** features a nitrogen atom instead of a carbon atom at the 6-position, which possesses a lone pair that blocks any nucleophilic addition to this position. Compound **3** is consequently a perfect model system to investigate whether such a nucleophilic activation is required for the deformylation, as mechanistically postulated (Figure 1A).

The nucleosides **2** and **3** were subsequently added at a concentration of 350 μM to the media of Neuro-2a, RBL-2H3, CHO-K1 cells for 72 hours (see the Supporting Information). During this time, the nucleosides are converted in vivo into the corresponding triphosphates and then incorporated into the genome of the dividing cells. Initial studies in which we fed the nucleosides individually allowed us to determine that neither compound decreases cell viability up to a concentration of 400 μM, thus the experiments were conducted below the toxicity level. In addition, we tested **2** and **3** at 350 μM on E14 TDG +/− and −/− mouse embryonic stem cells (mESCs) under a three-day priming process with C/R media. This system allowed us to exclude the BER pathway, leading to a detectable and quantifiable accumulation of natural fdC (see the Supporting Information). After three days, the cells were harvested and lysed, and the genomic DNA was extracted using an optimised protocol (see the Supporting Information). This was followed by an enzymatic digestion of gDNA to single nucleosides and analysed according to a method that we reported recently in detail.[21] The obtained nucleoside mixture containing mostly the canonical nucleosides dA, dC, dG, and dT, plus the non-canonical nucleosides mdC, hmdC, and fdC, as well as the incorporated molecules F-fdC and a-fdC and their potential downstream products (F-dC, F-mdC, a-dC, a-mdC). Nucleosides were separated by ultra-HPL chromatography and characterized by coupling of the UHPLC system to a triple-quadrupole mass spectrometer. For exact quantification of the nucleosides by isotope dilution, isotopically labelled standards of F-fdC and of the product F-dC were spiked into the analysis mixture as internal standards (see the Supporting Information). To enable exact quantification, calibration curves using these standards were determined (see the Supporting Information). Quantification was performed in the linear region.

During the analysis, we noted that an unusually low amount of a-fdC (**3**) was detected because it showed a broad elution profile with very low intensity (Figure 2B). All attempts to sharpen the elution profile in order to gain sensitivity failed. NMR analysis of compound **3** showed the reason for broad elution profile (see the Supporting Information). Due to the additional electron-withdrawing in-ring nitrogen atom, compound **3** exists partially as its hydrate in aqueous solution (20%, see the Supporting Information).

 © 2020 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

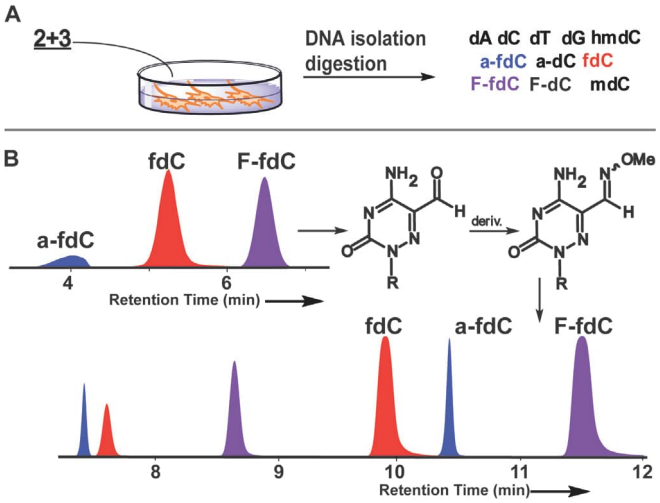


Figure 2. A) Overview of the experimental steps with the feeding and analysis. B) Analysis scheme and the reaction of a-fdC with methoxyamine to block hydrate formation and of a typical UHPL-chromatogram before (C-8 column) and after derivatization (C-18 column) for exact quantification. Peak splitting is due isomerization (blue peaks: a-fdC, red peaks: fdC, and purple peaks: F-fdC).

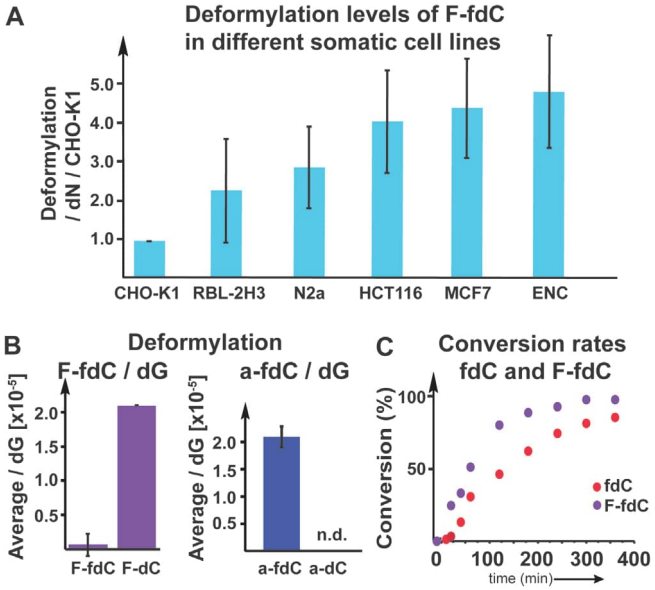


Figure 3. A) Deformylation data for F-fdC in different cell types, showing that F-fdC is deformylated in very different cells. Deformylation rate was calculated by the F-dC + F-mdC/dN per F-fdC/dN, then the values were normalized to the cell line with the lowest deformylation level (CHO-K1 = 1). B) The deformylation of F-fdC/dG and a-fdC/dG, showing the induced differences due to C6-carbon-to-nitrogen exchange. C) The bisulfite data show that the deformylation of fdC and F-fdC is comparable, thus showing that the 2′-F substitution has only a small accelerating effect, whereas the reaction of a-fdC could not be detected.

Although the ease of hydrate formation may foster deformylation, the hydrate/carbonyl equilibrium makes efficient detection of compound **3** basically impossible. In order to circumvent the problem, we started to derivatize a-fdC (**3**) before analysis with methoxyamine. Addition of CH_3ONH_2 (150 mM) to the digestion solution indeed provided the methoxyoxime of a-fdC in quantitative yield after just 15 min at 25 °C and pH 10. The naturally present fdC (**1**) and the probe molecule F-fdC (**2**), however, react as well, but unfortunately not quantitatively. To reduce impurities during MS measurements, we decided against using a catalyst for oxime formation. We therefore decided to analyse the digested DNA in two batches. The first one contained the digested untreated DNA to quantify all bases other than a-fdC. In the second batch, we treated the digested DNA with methoxyamine for a-fdC quantification. For quantification of the derivatized a-fdC, we constructed an external calibration curve (see the Supporting Information).

With this method in hand, we next quantified all nucleosides present in the genome of the cells treated with a mixture of **2** + **3**.

Figure 3 A shows that we indeed detected the fluorinated F-dC (**2**), thus confirming very efficient deformylation activity. We tested different cell types and found different levels of deformylation activity. But in all cases, the conversion of F-fdC into F-dC was clearly detectable. Most interesting is that we observed the highest deformylation activity in cells associated with neuronal properties. This is in line with neurons featuring the highest levels of hmdC and fdC. In contrast, Figure 3 B shows that for a-fdC (**3**), we were unable to detect any formation of the deformylated compound a-dC despite the high propensity of **3** to exist in the hydrated form, which is one prerequisite for efficient C–C bond cleavage. This result suggests that the ability to react with a nucleophile at the 6-position is also required in vivo for efficient deformylation.

In order to substantiate this result, we next performed in vitro studies with bisulfite. Bisulfite is a strong nucleophile that has been reported to cause deformylation of fdC by first attacking the C6 position, followed by conversion of the C5, C6-saturated fdC adduct into the bisulfite adduct, which then undergoes deformylation.[22] The deformylated product dC is then further converted into dU by the well-known bisulfite-induced deamination reaction of dC (see the Supporting Information). Indeed, when we reacted fdC with bisulfite, we observed efficient deformylation and deamination to dU. We then studied to what extent the reaction is influenced by the 2′-F atom present in F-fdC, in order to estimate whether the in vivo deformylation could be just the result of the 2′-F atom. Treatment of F-fdC with bisulfite also led to deformylation and deamination to F-dU, and indeed the reaction is a little faster compared to fdC (see Figure 3 C). Although the difference is measurable, it is in total rather small. With these data in hand, we can conclude that we may overestimate the amount of deformylation that can occur with fdC lacking the 2′-F atom. We can certainly exclude that deformylation in vivo occurs only with F-fdC. It is unfortunate that we are unable to measure the direct deformylation of fdC because of the presence of efficient BER processes. A TDG −/− cell line showed a huge increase in fdC compared to the TDG +/−, whereas a-fdC and F-fdC stay constant, thus showing that these compounds are indeed not repaired by the

 © 2020 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

TDG protein (see the Supporting Information). The bisulfite studies, however, show that the F-fdC compound is not a perfect but sufficient reporter of this C–C bond cleavage. Treatment of a-fdC (**3**) with bisulfite did not provide the deformylated product a-dC under any circumstances, showing that the inability to react with a nucleophile at the 6-position totally blocks the C–C bond cleavage. We can therefore conclude that the deformylation of fdC during active demethylation requires oxidation of mdC to fdC. fdC can undergo a direct C–C bond cleavage to dC, but this reaction requires a helper nucleophile to attack the C6-position, which is blocked in the case of a-fdC by the lone pair introduced by the C6-carbon-to-nitrogen exchange. While the chemistry that allows the transformation of fdC into dC is now elucidated, we next need to find the nucleophiles that perform the reaction in vivo.

Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) GRK2338 (Project ID 321812289), SFB1309 (PID 325871075), SFB1361 (PID 393547839) and SPP1784 (PID 255344185). This project has received additional funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n° EPiR 741912) and through a H2020 Marie Skłodowska-Curie Action (LightDyNAmics, 765866).

Conflict of interest

The authors declare no conflict of interest.

Keywords: demethylation · DNA modifications · epigenetics · formylcytidine

How to cite: *Angew. Chem. Int. Ed.* **2020**, *59*, 5591–5594
Angew. Chem. **2020**, *132*, 5639–5643

[1] M. Wagner, J. Steinbacher, T. F. J. Kraus, S. Michalakis, B. Hackner, T. Pfaffeneder, A. Perera, M. Müller, A. Giese, H. A. Kretzschmar, T. Carell, *Angew. Chem. Int. Ed.* **2015**, *54*, 12511–12514; *Angew. Chem.* **2015**, *127*, 12691–12695.

[2] S. Liu, J. Wang, Y. Su, C. Guerrero, Y. Zeng, D. Mitra, P. J. Brooks, D. E. Fisher, H. Song, Y. Wang, *Nucleic Acids Res.* **2013**, *41*, 6421–6429.

[3] C. X. Song, K. E. Szulwach, Q. Dai, Y. Fu, S. Q. Mao, L. Lin, C. Street, Y. Li, M. Poidevin, H. Wu, J. Gao, P. Liu, L. Li, G. L. Xu, P. Jin, C. He, *Cell* **2013**, *153*, 678–691.

[4] T. Pfaffeneder, B. Hackner, M. Truss, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 7008–7012; *Angew. Chem.* **2011**, *123*, 7146–7150.

[5] M. Bachman, S. Uribe-Lewis, X. Yang, H. E. Burgess, M. Iurlaro, W. Reik, A. Murrell, S. Balasubramanian, *Nat. Chem. Biol.* **2015**, *11*, 555–557.

[6] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929–930.

[7] D. Q. Shi, I. Ali, J. Tang, W. C. Yang, *Front. Genet.* **2017**, *8*, 100.

[8] M. Münzel, D. Globisch, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 6460–6468; *Angew. Chem.* **2011**, *123*, 6588–6596.

[9] C. G. Spruijt, F. Gnerlich, A. H. Smits, T. Pfaffeneder, P. W. T. C. Jansen, C. Bauer, M. Münzel, M. Wagner, M. Müller, F. Khan, H. C. Eberl, A. Mensinga, A. B. Brinkman, K. Lephikov, U. Müller, J. Walter, R. Boelens, H. Van Ingen, H. Leonhardt, T. Carell, M. Vermeulen, *Cell* **2013**, *152*, 1146–1159.

[10] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300–1303.

[11] F. Neri, D. Incarnato, A. Krepelova, S. Rapelli, F. Anselmi, C. Parlato, C. Medana, F. DalBello, S. Oliviero, *Cell Rep.* **2015**, *10*, 674–683.

[12] M. Su, A. Kirchner, S. Stazzoni, M. Müller, M. Wagner, A. Schröder, T. Carell, *Angew. Chem. Int. Ed.* **2016**, *55*, 11797–11800; *Angew. Chem.* **2016**, *128*, 11974–11978.

[13] T. Fu, L. Liu, Q. L. Yang, Y. Wang, P. Xu, L. Zhang, S. Liu, Q. Dai, Q. Ji, G. L. Xu, C. He, C. Luo, L. Zhang, *Chem. Sci.* **2019**, *10*, 7407–7417.

[14] R. Rahimoff, O. Kosmatchev, A. Kirchner, T. Pfaffeneder, F. Spada, V. Brantl, M. Müller, T. Carell, *J. Am. Chem. Soc.* **2017**, *139*, 10359–10364.

[15] S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Höfner, K. T. Wanner, C. Ochsenfeld, T. Carell, *J. Am. Chem. Soc.* **2013**, *135*, 14593–14599.

[16] S. Schiesser, B. Hackner, T. Pfaffeneder, M. Müller, C. Hagemeier, M. Truss, T. Carell, *Angew. Chem. Int. Ed.* **2012**, *51*, 6516–6520; *Angew. Chem.* **2012**, *124*, 6622–6626.

[17] K. Iwan, R. Rahimoff, A. Kirchner, F. Spada, A. S. Schröder, O. Kosmatchev, S. Ferizaj, J. Steinbacher, E. Parsa, M. Müller, T. Carell, *Nat. Chem. Biol.* **2018**, *14*, 72–78.

[18] A. Jeltsch, *ChemBioChem* **2002**, *3*, 274–293.

[19] Q. Du, Z. Wang, V. L. Schramm, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 2916–2921.

[20] I. V. Alekseeva, A. S. Shalamai, V. S. Shalamai, V. P. Chernetski, *Ukr. Khim. Zh. (Ukr. Ed.)* **1976**, *42*, 398–401.

[21] F. R. Traube, S. Schiffers, K. Iwan, S. Kellner, F. Spada, M. Müller, T. Carell, *Nat. Protoc.* **2019**, *14*, 283–312.

[22] E. Kriukienė, Z. Liutkevičiūtėa, S. Klimašauskas, *Chem. Soc. Rev.* **2012**, *41*, 6916–6930.

Manuscript received: January 9, 2020
Accepted manuscript online: January 30, 2020
Version of record online: February 25, 2020

 © 2020 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

3.4 Biomimetic iron complex achieves TET enzyme reactivity

D. Schmidl*, N. Jonasson*, E. Korytiaková*, T. Carell, L. Daumann, *Angew. Chem. Int. Ed.* **2021**, *60,* 21457-21463.

*: Diese Autoren haben zu gleichen Teilen beigetragen

Prolog

Die aktive Demethylierung von dem wichtigsten epigenetischen Marker mdC verläuft entweder über eine passive Demethylierung oder über eine schrittweise Oxidation von mdC zu hmdC, fdC und cadC durch Tet Enzyme. Die oxidierten mdC Nukleobasen werden über die Tdg-vermittelte Basenexzisionsreparatur oder eine direkte Defunktionalisierung mittels Deformylierung von fdC und Decarboxylierung von cadC demodifiziert. Es wurde postuliert, dass die Tet-vermittelte Oxidation von den methylierten Cytidinmodifikationen über einen aktiven Eisen(IV)-Oxo Intermediat abläuft.[179] *Jonasson et al*.[180-182] haben auf der Ebene der Methylcytosinbase gezeigt, dass der synthetische Eisen(IV)-Oxo Komplex $[Fe^{IV}(O)(Py_5Me_2H)]^{2+}$, (Fe-**1**) ein biomimetisches Modell für Tet-Enzyme darstellt und die Reaktivität des aktiven Zentrums der Tet-Enzyme modelliert. Im Rahmen dieser Arbeit wurde in einer Reihe von HPLC- und MS-Experimenten gezeigt, dass bei der Oxidation von Nukleosiden und kurzen mdC haltigen Oligonukleotiden mit Fe-**1** die natürliche Reaktivität der Tet-Enzyme erreicht werden konnte. Der Modellkomplex oxidiert im Kontext eines mdC enthaltenden Oligonukleotids die 5-Methylgruppe von mdC mit extrem hoher Selektivität. Es wurde eine detaillierte Reaktivitätsstudie mit methylierten Cytidinderivaten und Oligonukleotiden durchgeführt um die möglichen Nebenreaktionen beim Einsatz von Fe-**1** zu untersuchen. Als einzige relevante Nebenreaktion wurde die Oxidation des anomeren Zentrums des Nukleosids indentifiziert. Jedoch verlief die Oxidation der C-5-Methylgruppe des mdC mit Fe-**1** sehr selektiv und Fe-**1** konnte dadurch als ein idealer biomimetischer Modellkomplex für Tet-Enzyme auf Nukleosid und Oligonukleotidebene gezeigt werden.

Autorenbeitrag

In dieser Arbeit war ich für den Verdau der Oligonukleotide, die ensprechende massenspektrometrische Analyse (mittels UHPLC-MS/MS, Orbitrap LC-MS und MALDI) der untersuchten Proben und die Datenauswertung zuständig. *David Schmidl* führte die Synthese der Nukleoside, die Eisen-Oxidationsexperimente auf Nukleosidebene sowie die

ensprechende HPLC Analyse durch. *Niko Jonasson* war für die Synthese des Eisenkomplexes sowie die Eisenoxidationsassays mit den Oligonukleotiden und deren Probenaufbereitung zuständig. Alle Experimente und erhaltene Daten wurden in enger Zusammenarbeit mit *David Schmidl*, *Niko Jonasson*, *Lena Daumann* und *Thomas Carell* geplant und diskutiert.

Lizenz

Kopie der Open Access Publikation auf Basis der Creative Commons Attribution-NonCommercial License (CC BY-NC 4.0), die den nicht-kommerziellen Abdruck des Artikels erlaubt.

Artikel einsehbar unter:

https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.202107277

Lizenz einsehbar unter:

https://creativecommons.org/licenses/by-nc/4.0/

Die *Supporting Information* dieser Veröffentlichung ist im Anhang (Kapitel 8.3) abgebidet.

Check for updates

How to cite: Angew. Chem. Int. Ed. 2021, 60, 21457–21463
International Edition: doi.org/10.1002/anie.202107277
German Edition: doi.org/10.1002/ange.202107277

Bioinorganic Chemistry

Biomimetic Iron Complex Achieves TET Enzyme Reactivity**

*David Schmidl[+], Niko S. W. Jonasson[+], Eva Korytiaková[+], Thomas Carell, and Lena J. Daumann**

Abstract: *The epigenetic marker 5-methyl-2′-deoxycytidine (5mdC) is the most prevalent modification to DNA. It is removed inter alia via an active demethylation pathway: oxidation by Ten-Eleven Translocation 5-methyl cytosine dioxygenase (TET) and subsequent removal via base excision repair or direct demodification. Recently, we have shown that the synthetic iron(IV)-oxo complex $[Fe^{IV}(O)(Py_5Me_2H)]^{2+}$ (**1**) can serve as a biomimetic model for TET by oxidizing the nucleobase 5-methyl cytosine (5mC) to its natural metabolites. In this work, we demonstrate that nucleosides and even short oligonucleotide strands can also serve as substrates, using a range of HPLC and MS techniques. We found that the 5-position of 5mC is oxidized preferably by **1**, with side reactions occurring only at the strand ends of the used oligonucleotides. A detailed study of the reactivity of **1** towards nucleosides confirms our results; that oxidation of the anomeric center (1′) is the most common side reaction.*

Introduction

5-methyl-2′-deoxycytidine (5mdC), an epigenetic marker, is the most common natural DNA modification.[1] The corresponding nucleobase, 5mC, is also referred to as the fifth base, in addition to the canonical bases A, T, C and G. 5mdC is generated in the mammalian genome primarily but not solely in CpG dinucleotides via methylation of cytosine by DNA methyl transferases.[1,2] Most of these CpG dinucleotides (up to 80%) are methylated, although they remain largely unmodified when located in CpG islands within promotor regions.[3] 5mdC levels are fairly stable making up to 5% of all cytosines in adult mammalian cells.[4] However, their fate is much more dynamic in embryonic stem cells during the early stages of development.[5,6] 5mdC levels are regulated either via dilution during cell division (referred to as passive demethylation) or via a stepwise oxidation mechanism where 5mdC is converted to 5-hydroxymethyl-2′-deoxy-cytidine (5hmdC), 5-formyl-2′-deoxy-cytidine (5fdC), and 5-carboxy-2′-deoxy-cytidine (5cadC) by Ten-eleven Translocation 5-methyl cytosine Dioxygenases (TET, active demethylation).[7–10] 5fdC and 5cadC are then removed either via base excision repair or via direct demodification.[9,11] TET belongs to the superfamily of non-heme iron(II)/α-ketoglutarate (α-KG) dependent enzymes. These enzymes possess a conserved active site that consists of the so-called facial triad (two histidine residues and one carboxylate-containing amino acid, in the case of TET: glutamate) that coordinates to an iron ion. The coordination sphere is completed by coordination of the co-factor α-KG, water, or oxygen.[12,13] During activation, α-KG is converted to succinate while an iron(IV)-oxo species (see Figure 1 left) is generated.[14–16] This iron(IV)-oxo moiety is believed to be the active species that abstracts a hydrogen atom from the substrate, generating an iron(III)-hydroxo species and a carbon-centered radical. These then recombine to form the hydroxylated product.[17] Multiple iron(IV)-oxo complexes have been used as model systems to mimic natural processes.[18–24] The iron(IV)-oxo complex utilized in this work ($[Fe^{IV}(O)(Py_5Me_2H)]^{2+}$ (**1**), Py_5Me_2H = 2,6-bis(1,1-bis(2-pyridyl)ethyl)pyridine, counterions: F^-, OH^-, NO_3^-) was first described by Chang et al. in 2015.[25] **1** is soluble in water, air-stable, and does not decompose within several hours at room temperature.[25] We have recently demonstrated that **1** is capable of mimicking the behavior of TET enzymes by oxidizing the nucleobase 5mC to 5hmC, 5fC and 5caC. By comparing the reaction rates of **1** with those published for TET2,[26] we were able to conclude that the second coordination sphere of TET enzymes is one important factor responsible for its substrate preference. Furthermore, GC/

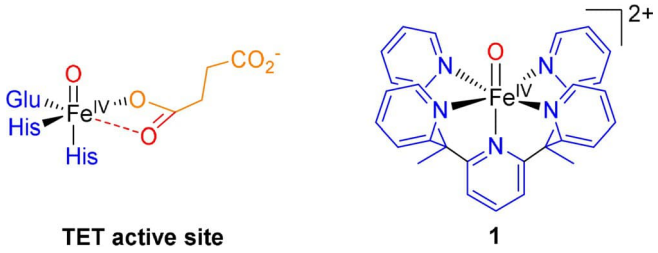


Figure 1. The active site of TET enzymes (left), highlighted in blue is the facial triad, in red are the two oxygen atoms that originate from molecular dioxygen and in orange is the remainder of the α-KG co-substrate.[12] Right: Biomimetic iron(IV)-oxo complex **1** used in this study with the active oxo moiety shown in red.

[*] D. Schmidl,[+] N. S. W. Jonasson,[+] E. Korytiaková,[+] T. Carell, L. J. Daumann
Department Chemie, Ludwig-Maximilians-University München
Butenandtstr. 5–13, Haus D, München (Germany)
E-mail: lena.daumann@lmu.de
[+] These authors contributed equally to this work.

[**] Ten-eleven translocation 5-methylcytosine dioxygenase.

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/anie.202107277.

© 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

MS and UV-vis studies confirmed hydrogen atom abstraction to be the rate determining step.[27]

The selective oxidation or reduction of cytosine derivatives in DNA is a highly desired transformation for in vitro processes, such as the sequencing of epigenetic bases. Whilst several chemical methods have been published for the oxidized bases,[28–32] the controlled conversion of 5mC still poses a challenge due to the low reactivity of the methyl group and its chemical similarity to the sugar scaffold. Selective oxidation is even more desirable as the current gold standard for single-base 5mC sequencing, the bisulfite method,[33] is destructive to the DNA sample[34,35] and is not able to intrinsically distinguish 5mC/5hmC and C/5fC/5caC.[28,36–39] Hence, the reactivity of TET enzymes presents a useful tool demonstrated by TAB-Seq[40] (TET-assisted bisulfite sequencing) for 5hmC and TAPS[41] (TET-assisted pyridine-borane sequencing) for 5mC and 5hmC. Nevertheless, the handling of TET enzymes is non-trivial and one major drawback of these methods is the large amount of enzyme required (sometimes more than stoichiometric amounts[40]). Moreover, the substrate specificity in terms of sequence bias/sequence preference and divergent reactivity against ssDNA and dsDNA might limit the utility of enzyme-based approaches. This technological example illustrates that a chemical strategy mimicking the reactivity of TET enzymes might benefit in vitro, and potentially even in vivo, applications for selective oxidation of 5mdC. In the present work we show that the substrate scope of **1** includes nucleosides and oligonucleotides. We show that **1** is capable of selectively oxidizing a 5mdC residue within a 10-mer oligonucleotide context. We also present a comprehensive study of the reactivity of **1** towards nucleosides including side reactions affecting the ribose unit.

Results and Discussion

After the successful demonstration of the oxidation of the nucleobase 5mC by **1**, we wanted to take a step towards nucleosides and actual DNA oligonucleotides as a substrate to evaluate the potential of **1** for biomimetic-TET-based sequencing methods. We therefore studied the reactivity of **1** towards a short oligomer (10-mer) containing a 5mdC residue (see Figure 2).

Oligonucleotide Substrates

We chose an oligonucleotide sequence which is composed of deoxyadenosines, deoxycytidines, deoxyuridines and deoxyguanosines, as these nucleobases do not contain aliphatic hydrocarbons that are prone to be attacked by **1**. Deoxyuridines were included instead of thymidines, as the methyl group of thymine is, of course, a possible target for oxidation by **1**. As 5mdC residues occur in nature, mostly in CpG dinucleotides, we included this into our oligonucleotide design and therefore chose the following sequence (5′->3′): HO-CCUUAACC**5mdC**G-OH (**O**). When **O** was exposed to **1** (1.0 or 4.0 equiv) we detected significant amounts of the

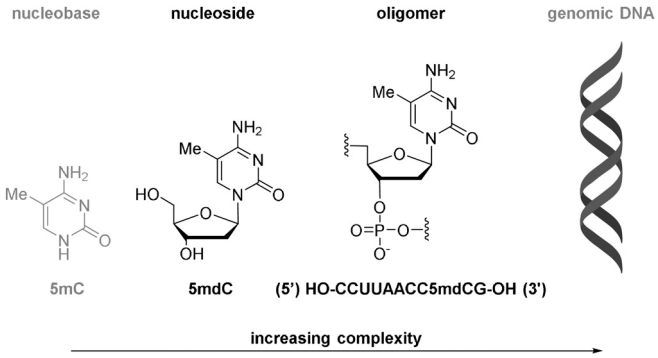


Figure 2. The increasing complexity of the substrate: from nucleobase via nucleoside and oligonucleotide to genomic DNA. In this work we present results from reactions containing nucleosides and oligonucleotides.

oxidized derivatives of 5mdC (For experimental details please see supporting information sections 1.6, 1.8 and 3). Figure 3 shows the amounts detected for 5mdC, 5hmdC, 5fdC and 5cadC in two samples containing both **O** and **1** (1 equiv **1**, 4 equiv **1**) as well as two control samples (**O** control, **1** control) that only contain the specified component. The common oxidative DNA lesion 8-oxo-dG was not observed (not shown, see Figure S27).

As observed in Figure 3, higher amounts of iron complex **1** generally result in higher levels of 5hmdC, 5fdC and 5cadC. This is pronounced least for 5hmdC and most for 5cadC, which points to a sequential oxidation process. It should be noted, that levels for 5hmdC and 5fdC are 10-fold higher than

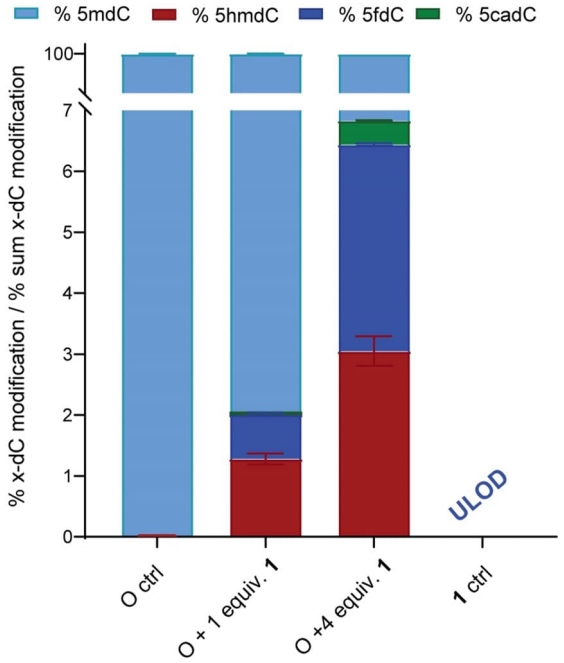


Figure 3. Detected amounts of 5hmdC, 5fdC, 5cadC when **O** was reacted with 1.0 equiv of **1** (O + 1 equiv **1**) or 4.0 equiv of **1** (O + 4 equiv **1**) compared to control measurements containing only **O** (**O** ctrl) or **1** (**1** ctrl) that were treated exactly as the samples. Conditions: [**O**] = 0.25 mM, [**1**] = 0.25 mM/1.0 mM, H_2O, T = 22 °C, reaction time = 30 min. ULOD refers to under level of detection.

 © 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

for 5cadC. This circumstance indicates both a stepwise oxidation process and a higher reactivity of 5hmdC vs. 5fdC (see also section on nucleoside substrates).

8-oxo-dG is a common oxidative DNA lesion that can cause base mismatching.[42,43] The formation of 8-oxo-dG is a potential side-reaction in oxidative treatment of DNA, as dG presents the lowest oxidation potential of the canonical bases.[44] However, this was not observed in the case of **O**: 8-oxo-dG levels are under the level of detection (ULOD) for all samples including control samples (Figure S27). Thus, we conclude that 8-oxo-dG is not formed upon treatment of oligonucleotides with **1**.

Additionally, we tracked the time-resolved development of 5mdC, 5hmdC, 5fdC, and 5cadC levels in the reaction of **1** with **O**. 5mdC levels decrease (Figure S28) whereas levels of 5hmdC and 5fdC increase sharply immediately after addition of **1** (Figure 4). 5cadC levels also increase but to a much smaller extent. It is noteworthy that the final concentrations observed in these experiments match the concentrations observed in the experiments described above (Figure 3): 5hmdC and 5fdC are about 10-fold more abundant than 5cadC. This shows that the method is reproducible. The circumstance where we detected different ratios for 5hmdC/5cadC and 5fdC/5cadC (see Figure S29) when using different amounts of **1**, hints that the amount of **1** added can be used to stir the product formation of the reaction. Additionally, the development of the product ratios over time indicates that the reaction time may provide a similar method of influencing the product distribution for possible sequencing applications (Figure S29). The concentration of the substrate (0.125–0.25 mM) in the experiments discussed above was set at these levels so that the results are comparable with our previous studies and the reactions with nucleosides as substrates described below. For a biochemical application, the substrate concentrations need to be adjusted accordingly.[27]

Further studies with 5mdC residues within larger oligomeric structures, double-stranded substrates, and at lower substrate concentrations are necessary, but our findings give a first insight into the usefulness of **1** in the controlled oxidation of methylated or oxomethylated DNA samples, for example, for sequencing applications. In contrast to the previously studied nucleobases, oligonucleotides present more than one point of possible attack: the methyl group on the nucleobase itself or the carbohydrate (deoxyribose) backbone can be subject to C–H activation. We therefore screened for possible side products (see Section 3.6 in the SI). Reaction mixtures were analyzed prior to enzymatic digestions for any of the target nucleosides (dG, 8-oxo-dG, dU, dC, 5mdC, 5hmdC, 5fdC, 5cadC) using triple QQQ mass spectrometry, however, detected levels were found to be below the limit of detection.

MALDI MS measurements show that the strand ends are oxidized upon exposure of **O** to **1**, but no internal strand-breaks occur (Figures S30–S39). We did detect the loss of ≈95 *m/z* and ≈135 *m/z* which would correspond to loss of a cytosine or guanine nucleobase fragment while simultaneously an oxygen atom is gained (Figures S32 and S34, Scheme S4). A likely mechanism is the oxidation of the 1′ position of the terminal nucleotides (deoxycytidine on the 5′ end, deoxyguanosine on the 3′ end) with subsequent loss of the nucleobase and formation of a lactone (Scheme 1 A). We also observed loss of ≈291 *m/z* (Figures S31, S33), which was

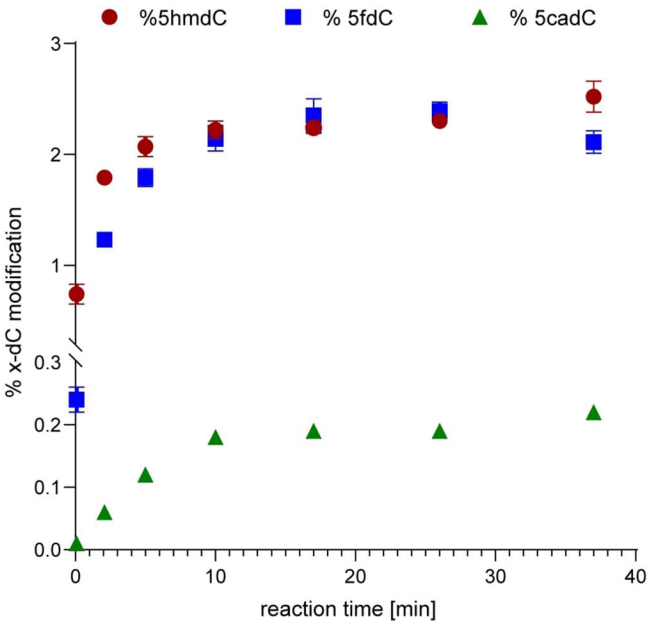
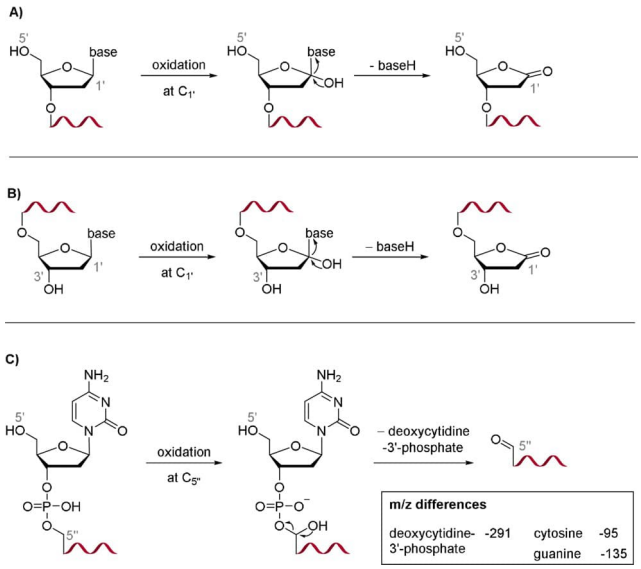


Figure 4. Time-resolved monitoring of the reaction of **O** with **1**: detected amounts of 5hmdC, 5fdC, 5cadC after certain time intervals. Conditions: [**O**] = 0.125 mM, [**1**] = 0.5 mM, H_2O, T = 22 °C. Data points are averaged from two replicates, each sample was measured in three technical replicates. For a separated representation of these data, including 5mdC levels, see Figure S28.



Scheme 1. A) Proposed mechanism for the loss of a nucleobase via oxidation of the 1′ position of the terminal deoxyribose unit on the 5′ end. B) Proposed mechanism for the loss of a nucleobase via oxidation of the 1′ position of the terminal deoxyribose unit on the 3′ end. C) Proposed mechanism for the loss of deoxycytidine-3′-phosphate via oxidation of the 5′-site of the second-to-last deoxyribose unit on the 5′ end of the strand (see Figures S30–S39 and Scheme S4 for a more detailed side product analysis).

 © 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

assigned to the oxidation of the 5′ position of the second-to last ribose unit on the 5′ end of the strand resulting in the loss of a cytidine-3′-phosphate fragment (Scheme 1B). We performed the same experiments with a control strand that contains a dC residue instead of 5mdC, and the same control strand that in addition possesses a phosphate cap on the 5′ end (for sequence details see Table S4). We observed significantly more side reactions in both cases (see Figures S36–S39). This confirms that oxidation at the 5-position of cytosine in the case of **O** is preferred over the side-reactions.

Interestingly, in all samples we observed either the loss of two cytosine fragments or the loss of one guanine and one cytosine fragment but never any interior strand breaks. This supports our interpretation that only the terminal positions on the oligonucleotide are subject to side reactions, demonstrating the highly specific reactivity of **1**.

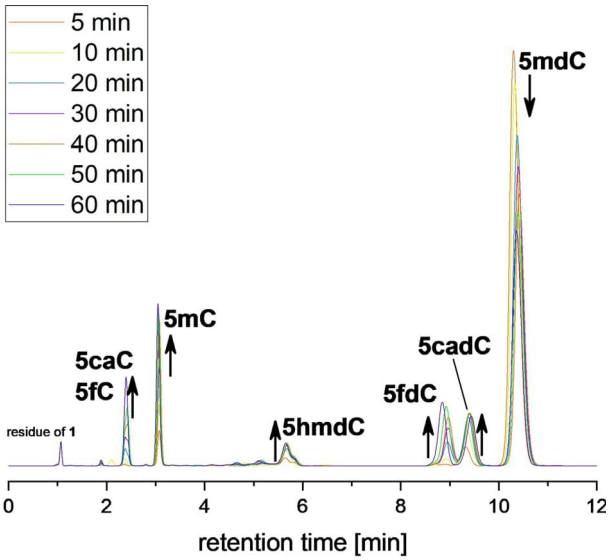
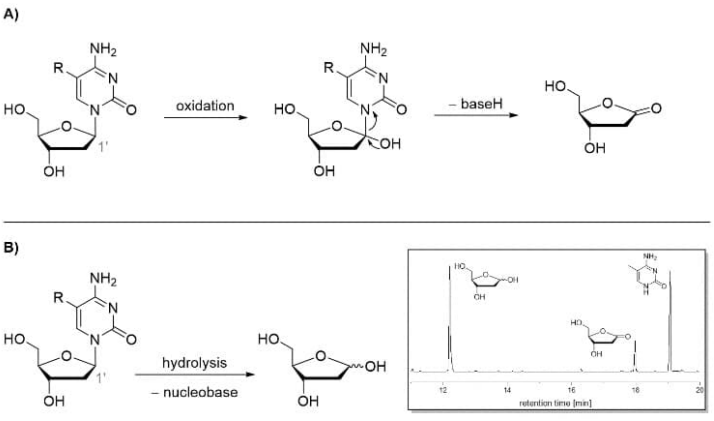


Figure 5. Excerpt of a representative HPLC trace of the reaction of 5mdC with **1**. Conditions: [5mdC] = 1 mM, [**1**] = 5 mM, H_2O, T = 22 °C. For a more detailed assignment see Figures S2–S10.

Nucleoside Substrates

To obtain a clearer picture of the competing reactivity of the potential oxidation sites in oligonucleotides, including confirmation of the above proposed side reaction pathways, we analyzed the reactivity of **1** towards the nucleoside substrates 5mdC and 5hmdC. Overall, we found that while the 1′ position on the deoxyribose unit is indeed oxidized forming the 2′-deoxyribono-γ-lactone, the 5-position on the cytosine base is oxidized faster yielding mostly the 5-oxomethylated cytidine derivatives 5hmdC, 5fdC, and 5cadC. We first conducted a series of reactions between 5mdC and different equivalents of **1**, for either 15 min or 30 min, to estimate how much and where the nucleoside with its high number of potential reactive sites would be oxidized. A color change from green to orange was observed during these reactions, indicating consumption of the green iron(IV)-oxo species **1**. Independent from the reaction time and the amount of added oxidant, we observed the 5-oxidized 5mdC derivatives 5hmdC, 5fdC, and 5cadC as the main products (Figure 5, Figure S17).

The relative abundance of nucleoside products reflects a correlation between the oxidation kinetics of the cytosine 5-substituent with the respective C–H bond dissociation energies (BDEs, see Table S5), as determined in previous studies with **1** and the nucleobases: 5hmC reacts the fastest, followed by 5mC and lastly 5fC.[27] Interestingly, these results are in opposition to those reported for TET enzymes (TET2), for which the following reaction rate sequence was found: $k_{5mC} > k_{5hmC} > k_{5fC}$.[26] While the behavior of TET2 was attributed to an additional interaction of the DNA substrate with a secondary coordination sphere in the enzyme,[27] the observed nucleoside product distribution in the reaction with **1** is in line with the consensus oxidation mechanism proposing an initial hydrogen atom transfer from the substrate to the iron(IV)-oxo species as the rate-determining step.[17,27] The corresponding nucleobases were detected on lower levels while practically no further side products, e.g., through oxidation at various deoxyribose carbons, were present in the product mixtures. Due to their similar BDEs, the anomeric center at the 1′ position of the nucleosides (BDE: 87.7 kcal mol^{-1})[45] might react with a similar rate as the most rapidly oxidized hydroxymethyl group of 5hmC (BDE: 86.2 kcal mol^{-1}).[26] This would lead to oxidative hydrolysis of the nucleoside (Scheme 2A, also compare Scheme 1A and B), constituting one possible explanation for the detection of nucleobase products upon addition of **1** to 5mdC. To confirm $C_{1'}$-oxidation, we subjected one such reaction mixture to GC/MS analysis and found 2′-deoxyribono-γ-lactone, the expected $C_{1'}$-oxidation side product, next to all four nucleobases (see inset in Scheme 2). It should be noted that the bases can not only be formed by oxidation of the anomeric center, but also



Scheme 2. A) Proposed mechanism for the loss of a modified cytosine nucleobase fragment from a nucleoside substrate via oxidation of the 1′ position of deoxyribose unit resulting in the formation of 2′-deoxyribono-γ-lactone. B) Proposed, although unlikely, hydrolysis of a modified cytidine nucleoside. A representative GC-MS trace is shown, the signals that were assigned to deoxyribose, 2′-deoxyribono-γ-lactone, and 5mC are indicated (see Figures S18–S22 for a more detailed analysis).

 © 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

via "normal" hydrolysis of the *N*-glycosidic bond, which might occur both during the reaction or the work-up process. Indeed, we also detected high amounts of 2′-deoxyribose; however, this is not a suitable indication for hydrolysis happening during the reaction as the nucleosides hydrolyze during the derivatization process prior to GC analysis. Since this has generally only been reported to happen under very acidic conditions[46] or with the use of enzymes,[47] it appears unlikely to have occurred in this case. By filtering a mixture of the four nucleoside standards through silica followed by freeze-drying and HPLC analysis, we could rule out that the work-up contributed significantly to nucleoside hydrolysis. Further, the side product analysis for the oligonucleotide reaction (Scheme 1 C) suggests that $C_{5'}$-oxidation also occurs which appears feasible due to the similar BDE to the 5-formyl group (BDEs: 92.4 kcal mol^{-1},[45] 92.9 kcal mol^{-1},[26] respectively). Such products might correspond to the remaining peaks observed in the HPLC chromatogram, however, those are so small that we propose the corresponding oxidation reactions to happen only to a very low extent, leaving anomeric center oxidation as the most prominent proposed side reaction on a nucleoside substrate.

To obtain a more detailed image of the progress of the oxidation reaction, 5mdC was reacted with 5 equivalents of **1** and the reaction tracked over 70 min by taking samples every 5 min (Figure 6, Scheme S1). Whereas the amount of recovered 5mdC decreased throughout the course of the reaction, the 5hmdC concentration reached a steady state at as low as < 5 % of recovered products after 25 min, indicating that it is more rapidly oxidized than 5mdC. Accordingly, the amount of 5fdC formed increased fast to reach its steady-state concentration between 7–9 % of recovered products, equally after 20–25 min, which shows that it is less quickly consumed than 5hmdC. In line with the proposed stepwise oxidation mechanism, significant 5cadC levels (i.e., three oxidation steps taking place) are detectable after 10 min at the earliest. Also, it seems that 5cadC is accumulated over the entire 70 min without reaching a steady state concentration within this timeframe. Nevertheless, 5mC levels reach up to 16 % and 5caC levels even add up to 18 % of recovered products, indicating that anomeric center oxidation is a significant side reaction on the nucleoside level.

Interestingly, the reaction of 5mdC with **1** seems to proceed much slower than that of **1** with the oligonucleotide **O**: The nucleoside reaction proceeds throughout the monitored 70 min time frame (Figure 6) while the reaction between **O** and **1** appears to be complete after 10–12.5 min (Figure 4)—even though the concentration of 5mdC is 8-fold higher than that of **O** and 5 equivalents of **1** are used for the nucleoside compared to 4 equivalents on the oligonucleotide. An explanation for this could be electrostatic attraction between negatively charged **O** and the iron(IV)-oxo cation bringing the reaction partners into spatial proximity. Additionally, π-stacking interactions might be feasible between the nucleobases and the aromatic pyridyl-based ligand of **1**.

An analogous time-resolved experiment of the reaction between 5hmdC with **1** was conducted (Figures S23, S24, Scheme S2). According to BDEs and previous experiments,[26,27] the hydroxymethyl group should react considerably faster than the methyl group of 5mdC, eventually leading to much higher levels of 5cadC. Additionally, more information might be gained about the relative reactivity of the anomeric center with its BDE being very similar to the hydroxymethyl group. Indeed, 5hmdC was much more rapidly converted than 5mdC, with only 46 % recovered already after 5 min, while it became almost completely consumed over the full 70 min (about 2.5 % recovery). In parallel, 5fdC was very swiftly formed, reaching its maximum concentration after about 10 min (ca. 45 % recovery) after which its concentration decreased again due to the formation reaction slowing down (ca. 20 % recovery after 70 min). 5cadC, which was already detected after 5 min, accumulated rapidly to become the main component in the reaction mixture, amounting to 75 % of

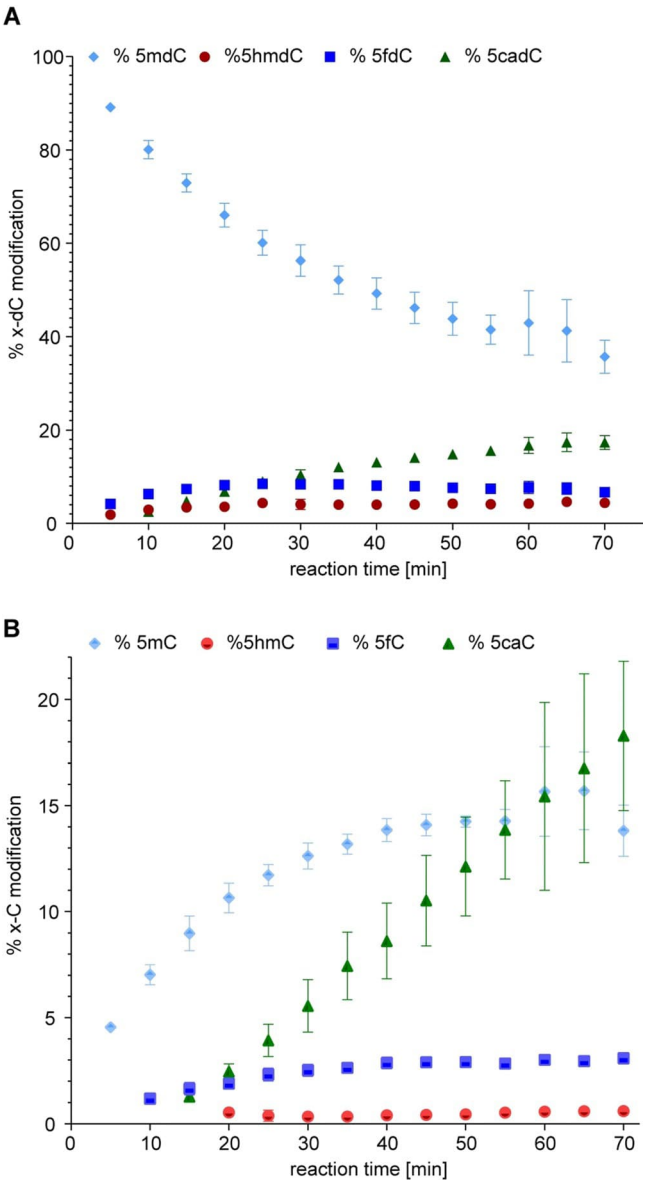


Figure 6. Time-resolved monitoring of the reaction products formed in the reaction of 5mdC with **1**: A) detected amounts of 5mdC, 5hmdC, 5fdC, 5cadC after certain time intervals; B) detected amounts of 5mC, 5hmC, 5fC and 5caC. Conditions: [5mdC] = 1.0 mM, [**1**] = 5.0 mM, H_2O, T = 22 °C. Data points are averaged from three replicates.

 © 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

recovered products after 70 min. Remarkably, the amount of nucleosides recovered remained at high proportions of the products recovered in total throughout the entire course of the reaction, again indicating that $C_{1'}$ oxidation was significantly slower than the oxidation of the respective 5-substituents.

A further control reaction was performed with 5-[D_3]-mdC (D_3-5mdC, for synthesis procedure see section 4 SI) as substrate for **1** (Figures S25, S26, Scheme S3), which should be oxidized very slowly at the 5-substituent and more preferably at $C_{1'}$ due to an expected kinetic isotope effect like we had previously observed for the nucleobases.[27] Indeed, only very low levels (under limit of quantification) of both 5-oxidised nucleosides and nucleobases were found even after 70 min. This kinetic isotope effect added more evidence for the consensus radical mechanism involving hydrogen atom abstraction as the rate-limiting step. While the amount of starting material decreased very slowly, D_3-5mC accumulated to about 21 % in the reaction mixture after 65 min. At the same time, the reaction solution retained a green color rather than the orange appearance it adopted with 5mdC and 5hmdC as substrates, both suggesting low consumption of **1** when the reaction with the cytosine 5-substituent is markedly slowed down.

The experiments with **1** and different nucleosides as substrates demonstrate conclusively that the most reactive position on the nucleoside is the 5-substituent of the nucleobase, modulated by its oxidation state. The anomeric center, on the other hand, appears less prone to oxidation by **1**, despite its lower BDE compared to the methyl group in 5mC or the formyl group in 5fC. This indicates that the $C_{1'}$ hydrogen atom is hindered by steric interactions while the base modification is easily accessible. The results are in line with the observations discussed above for oligonucleotide substrates, where elimination reactions caused by oxidation of the sugar backbone—which is significantly more sterically demanding than a single nucleoside sugar—occurred only marginally and, if, only at the terminal nucleotides.

Conclusion

We have provided clear evidence that the fully-synthetic biomimetic iron(IV)-oxo complex **1** achieved TET enzyme reactivity: the epigenetic marker 5mdC is oxidized to the corresponding TET metabolites 5hmdC, 5fdC, and 5cadC within an oligonucleotide context (albeit at relatively large concentrations of 0.125–0.25 mM). We monitored the product distribution for different amounts of **1** and its development over time. In these reactions, we observed few side products showing the high specificity of **1** toward oxidation of the 5-position on the cytosine moiety. The side reaction we observed, loss of a cytosine or guanine nucleobase, occurs only at the ends of the 10-mer oligonucleotide substrate. We postulate that oxidation at the 1′ position of deoxyribose forms a lactone and releases the nucleobase. This is supported by our studies of the reactivity of **1** towards nucleoside substrates 5mdC and 5hmdC. Here, we observed oxidation of the 5-position on the cytosine moiety to the expected derivatives (5hmdC/5fdC/5cadC). In addition, we found that 1′ oxidation occurs very slowly, producing the corresponding nucleobases and 2′-deoxyribono-γ-lactone, confirming our hypothesis concerning the oligonucleotide side reactions. Further studies should include longer and more complex oligonucleotides and double stranded substrates. In addition, the influence of the amount of **1** and the reaction time on product composition needs to be investigated as a potential tool for epigenetic sequencing.

Acknowledgements

A grant from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—SFB 1309-325871075 is gratefully acknowledged. NSWJ thankfully acknowledges financial support by a scholarship of the Studienstiftung des Deutschen Volkes. NSWJ also thanks Juliane Kamlah and Doreen Reuters neé Kremer for their contributions and Prof. Stefanie Kellner for helpful discussions. Open access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: 5-methyl cytosine · bioinorganic chemistry · DNA methylation · enzyme models · iron(IV)-oxo · TET enzymes

[1] M. V. C. Greenberg, D. Bourc'his, *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 590–607.
[2] Y. He, J. R. Ecker, *Annu. Rev. Genomics Hum. Genet.* **2015**, *16*, 55–77.
[3] A. M. Deaton, A. Bird, *Genes Dev.* **2011**, *25*, 1010–1022.
[4] T. Carell, M. Q. Kurz, M. Müller, M. Rossa, F. Spada, *Angew. Chem. Int. Ed.* **2018**, *57*, 4296–4312; *Angew. Chem.* **2018**, *130*, 4377–4394.
[5] E. Habibi, A. B. Brinkman, J. Arand, L. I. Kroeze, H. H. D. Kerstens, F. Matarese, K. Lepikhov, M. Gut, I. Brun-Heath, N. C. Hubner, R. Benedetti, L. Altucci, J. H. Jansen, J. Walter, I. G. Gut, H. Marks, H. G. Stunnenberg, *Cell Stem Cell* **2013**, *13*, 360–369.
[6] Z. D. Smith, A. Meissner, *Nat. Rev. Genet.* **2013**, *14*, 204–220.
[7] M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930–935.
[8] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300–1303.
[9] Y.-F. He, B.-Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C.-X. Song, K. Zhang, C. He, G.-L. Xu, *Science* **2011**, *333*, 1303–1307.
[10] F. Spada, S. Schiffers, A. Kirchner, Y. Zhang, G. Arista, O. Kosmatchev, E. Korytiakova, R. Rahimoff, C. Ebert, T. Carell, *Nat. Chem. Biol.* **2020**, *16*, 1411–1419.
[11] A. Schön, E. Kaminska, F. Schelter, E. Ponkkonen, E. Korytiaková, S. Schiffers, T. Carell, *Angew. Chem. Int. Ed.* **2020**, *59*, 5591–5594; *Angew. Chem.* **2020**, *132*, 5639–5643.
[12] E. L. Hegg, L. Que, *Eur. J. Biochem.* **1997**, *250*, 625–629.
[13] L. Hu, Z. Li, J. Cheng, Q. Rao, W. Gong, M. Liu, Y. G. Shi, J. Zhu, P. Wang, Y. Xu, *Cell* **2013**, *155*, 1545–1555.

 © 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

[14] J. C. Price, E. W. Barr, B. Tirupati, J. M. Bollinger, C. Krebs, *Biochemistry* **2003**, *42*, 7497–7508.
[15] P. J. Riggs-Gelasco, J. C. Price, R. B. Guyer, J. H. Brehm, E. W. Barr, J. M. Bollinger, C. Krebs, *J. Am. Chem. Soc.* **2004**, *126*, 8108–8109.
[16] M. L. Neidig, C. D. Brown, K. M. Light, D. G. Fujimori, E. M. Nolan, J. C. Price, E. W. Barr, J. M. Bollinger, C. Krebs, C. T. Walsh, E. I. Solomon, *J. Am. Chem. Soc.* **2007**, *129*, 14224–14231.
[17] S. Kal, L. Que, *J. Biol. Inorg. Chem.* **2017**, *22*, 339–365.
[18] P. A. MacFaul, K. U. Ingold, D. D. M. Wayner, L. Que, *J. Am. Chem. Soc.* **1997**, *119*, 10594–10598.
[19] M. Costas, K. Chen, L. Que, *Coord. Chem. Rev.* **2000**, *200–202*, 517–544.
[20] C. R. Goldsmith, R. T. Jonas, T. D. P. Stack, *J. Am. Chem. Soc.* **2002**, *124*, 83–96.
[21] M. R. Bukowski, K. D. Koehntop, A. Stubna, E. L. Bominaar, J. A. Halfen, E. Münck, W. Nam, L. Que, *Science* **2005**, *310*, 1000–1002.
[22] A. N. Biswas, M. Puri, K. K. Meier, W. N. Oloo, G. T. Rohde, E. L. Bominaar, E. Münck, L. Que, *J. Am. Chem. Soc.* **2015**, *137*, 2428–2431.
[23] J. E. M. N. Klein, L. Que, *Encyclopedia of Inorganic and Bioinorganic Chemistry*, American Cancer Society, Atlanta, **2016**, pp. 1–22.
[24] R. J. Martinie, C. J. Pollock, M. L. Matthews, J. M. Bollinger, C. Krebs, A. Silakov, *Inorg. Chem.* **2017**, *56*, 13382–13389.
[25] T. Chantarojsiri, Y. Sun, J. R. Long, C. J. Chang, *Inorg. Chem.* **2015**, *54*, 5879–5887.
[26] L. Hu, J. Lu, J. Cheng, Q. Rao, Z. Li, H. Hou, Z. Lou, L. Zhang, W. Li, W. Gong, M. Liu, C. Sun, X. Yin, J. Li, X. Tan, P. Wang, Y. Wang, D. Fang, Q. Cui, P. Yang, C. He, H. Jiang, C. Luo, Y. Xu, *Nature* **2015**, *527*, 118–122.
[27] N. S. W. Jonasson, L. Daumann, *Chem. Eur. J.* **2019**, *25*, 12091–12097.
[28] M. J. Booth, M. R. Branco, G. Ficz, D. Oxley, F. Krueger, W. Reik, S. Balasubramanian, *Science* **2012**, *336*, 934–937.
[29] W. Mao, J. Hu, T. Hong, X. Xing, S. Wang, X. Chen, X. Zhou, *Org. Biomol. Chem.* **2013**, *11*, 3568–3572.
[30] M. J. Booth, G. Marsico, M. Bachman, D. Beraldi, S. Balasubramanian, *Nat. Chem.* **2014**, *6*, 435–440.
[31] B. Xia, D. Han, X. Lu, Z. Sun, A. Zhou, Q. Yin, H. Zeng, M. Liu, X. Jiang, W. Xie, C. He, C. Yi, *Nat. Methods* **2015**, *12*, 1047–1050.
[32] C. Zhu, Y. Gao, H. Guo, B. Xia, J. Song, X. Wu, H. Zeng, K. Kee, F. Tang, C. Yi, *Cell Stem Cell* **2017**, *20*, 720–731.e5.
[33] M. Frommer, L. E. McDonald, D. S. Millar, C. M. Collis, F. Watt, G. W. Grigg, P. L. Molloy, C. L. Paul, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 1827–1831.
[34] A. M. Raizis, F. Schmitt, J. P. Jost, *Anal. Biochem.* **1995**, *226*, 161–166.
[35] K. Tanaka, A. Okamoto, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 1912–1915.
[36] Y. Huang, W. A. Pastor, Y. Shen, M. Tahiliani, D. R. Liu, A. Rao, *PLoS One* **2010**, *5*, e8888.
[37] C. Nestor, A. Ruzov, R. Meehan, D. Dunican, *Biotechniques* **2010**, *48*, 317–319.
[38] S.-G. Jin, S. Kadam, G. P. Pfeifer, *Nucleic Acids Res.* **2010**, *38*, e125.
[39] M. Yu, G. C. Hon, K. E. Szulwach, C.-X. Song, L. Zhang, A. Kim, X. Li, Q. Dai, Y. Shen, B. Park, J.-H. Min, P. Jin, B. Ren, C. He, *Cell* **2012**, *149*, 1368–1380.
[40] M. Yu, G. C. Hon, K. E. Szulwach, C.-X. Song, P. Jin, B. Ren, C. He, *Nat. Protoc.* **2012**, *7*, 2159–2170.
[41] Y. Liu, P. Siejka-Zielińska, G. Velikova, Y. Bi, F. Yuan, M. Tomkova, C. Bai, L. Chen, B. Schuster-Böckler, C.-X. Song, *Nat. Biotechnol.* **2019**, *37*, 424–429.
[42] K. C. Cheng, D. S. Cahill, H. Kasai, S. Nishimura, L. A. Loeb, *J. Biol. Chem.* **1992**, *267*, 166–172.
[43] S. Kanvah, J. Joseph, G. B. Schuster, R. N. Barnett, C. L. Cleveland, U. Landman, *Acc. Chem. Res.* **2010**, *43*, 280–287.
[44] F. Boussicault, M. Robert, *Chem. Rev.* **2008**, *108*, 2622–2645.
[45] S. Steenken, S. V. Jovanovic, L. P. Candeias, J. Reynisson, *Chem. Eur. J.* **2001**, *7*, 2829–2833.
[46] M. Sako, T. Kihara, H. Kawada, K. Hirota, *J. Org. Chem.* **1999**, *64*, 9722–9723.
[47] V. A. Stepchenko, F. Seela, R. S. Esipov, A. I. Miroshnikov, Y. A. Sokolov, I. A. Mikhailopulo, *Synlett* **2012**, *23*, 1541–1545.

Manuscript received: May 31, 2021
Revised manuscript received: June 23, 2021
Accepted manuscript online: June 28, 2021
Version of record online: August 20, 2021

 © 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

3.5 Active turnover of genomic methylcytosine in pluripotent cells

F. Spada, S. Schiffers, A. Kirchner, Y.-Q. Zhang, G. Arista, O. Kosmatchev, E. Korytiaková, R. Rahimoff, C. Ebert, T. Carell, *Nat. Chem. Biol.* **2020**,*16*, 1411–1419.

Prolog

Die Potenz einer Zelle wird durch die Dynamik der epigenetischen Prozesse bestimmt und reguliert. Es ist jedoch unklar, zu welchem Anteil die DNA-Methylierung und Demethylierung zu diesem regulatorischen Prozess beitragen. Die Demethylierung der DNA läuft passiv, durch DNA-Synthese, und aktiv, durch Tet abhängige Oxidationsprozesse, ab. Da eine Vielzahl an Prozessen zur Entfernung der Methylierung führen kann, ist die dynamische Demethylierung ein komplexer Prozess. Die Tet-oxidationsabhängige Demethylierung verläuft über Tdg-vermittelte Basenexzisionsreparatur und/oder einen direkten C-C-Bindungsbruch der oxidierten mdC Modifikationen. Welche Art an Demethylierung in welchen Anteilen zur epigenetischen Regulierung beiträgt ist bisher jedoch eine unbeantwortete Fragestellung. Im Rahmen dieser Arbeit wurden metabolische Markierungsexperimente in einer Reihe von pluripotenten Zelltypen der Maus mit stabilen Isotopen durchgeführt. Mit Hilfe von Massenspektrometrie wurden Quantifizierungsstudien von mdC und dessen Oxidationsprodukte hmdC, fdC und cadC auf globaler Ebene in den untersuchten Zellen durchgeführt. Die globalen Level von mdC und dessen Tet Oxidationsprodukte wurden in unterschiedlichen Stadien der Zellpluripotenz quantifiziert und verglichen. Es konnte gezeigt werden, dass bei pluripotenten Zellen nach dem *priming* 3-6% des genomischen mdC nach der Oxidation zu hmdC, fdC und cadC aus dem Genom entfernt wurde. Außerdem konnte eine oxidationsunabhängige mdC Demethylierung beobachtet werden, die mit DNA-Reparaturprozessen zusammenhängt. Ein kleinerer Teil des genomischen mdC wurde durch die Spaltung der Phosphodiester Bindung in das löslichen Nukleosid Pool freigesetzt und hier zu dT deaminiert. Im Anschluss wurde dT ins Genom wieder eingebaut. Diese Studie zeigt, dass in pluripotenten Zellen sowohl die Tet-oxidationsabhängigen als auch die oxidationsunabhängigen Prozesse für die dynamische DNA Demethylierung eine Rolle spielen.

Autorenbeitrag

Für diese Arbeit habe ich isotopenmarkierte Nukleosidderivate, die als interne Standards für massenspektrometrische Quantifizierung von Nukleosid Metaboliten in genomischer DNA eingesetzt wurden, synthetisiert. Der Hauptteil der Arbeit dieser Studie wurde von *Fabio Spada*

durchgeführt. *Sarah Schiffers* hat massenspektrometrische Methoden für die DNA und Nukleosid Pool Analyse entwicklet und die DNA-Proben verdaut und mittels Massenspektrometrie analysiert.

Lizenz

Kopie der Publikation mit Erlaubnis des Verlags; Copyright 2021 *Nature Publishing Group*.

DOI: 10.1038/s41589-020-0621-y

Artikel einsehbar unter:

https://www.nature.com/articles/s41589-020-0621-y

Die *Supporting Information* dieser Veröffentlichung ist einsehbar unter:

https://static-content.springer.com/esm/art%3A10.1038%2Fs41589-020-0621-y/MediaObjects/41589_2020_621_MOESM19_ESM.pdf

nature chemical biology

ARTICLES

https://doi.org/10.1038/s41589-020-0621-y

Check for updates

Active turnover of genomic methylcytosine in pluripotent cells

Fabio Spada[1]✉, Sarah Schiffers[1,3], Angie Kirchner[1,4], Yingqian Zhang[1,2], Gautier Arista[1], Olesea Kosmatchev[1], Eva Korytiakova[1], René Rahimoff[1,5], Charlotte Ebert[1] and Thomas Carell[1]✉

Epigenetic plasticity underpins cell potency, but the extent to which active turnover of DNA methylation contributes to such plasticity is not known, and the underlying pathways are poorly understood. Here we use metabolic labeling with stable isotopes and mass spectrometry to quantitatively address the global turnover of genomic 5-methyl-2′-deoxycytidine (mdC), 5-hydroxymethyl-2′-deoxycytidine (hmdC) and 5-formyl-2′-deoxycytidine (fdC) across mouse pluripotent cell states. High rates of mdC/hmdC oxidation and fdC turnover characterize a formative-like pluripotent state. In primed pluripotent cells, the global mdC turnover rate is about 3–6% faster than can be explained by passive dilution through DNA synthesis. While this active component is largely dependent on ten-eleven translocation (Tet)-mediated mdC oxidation, we unveil additional oxidation-independent mdC turnover, possibly through DNA repair. This process accelerates upon acquisition of primed pluripotency and returns to low levels in lineage-committed cells. Thus, in pluripotent cells, active mdC turnover involves both mdC oxidation-dependent and oxidation-independent processes.

Methylation at position five of 2′-deoxycytidine (dC) is the most prominent nucleobase modification in higher eukaryotic genomes[1]. In mammals, genomic dC is methylated to mdC by the DNA (cytosine-5)-methyltransferases DNMT1, DNMT3a and DNMT3b, mostly but not exclusively in the context of CpG dinucleotides[1,2]. These enzymes differentially contribute to establishing and maintaining genomic mdC patterns across cellular and organismal generations. Genomic mdC is further modified by iterative oxidation to hmdC, fdC and 5-carboxyl-dC (cadC) through the action of Tet dioxygenases[3–5]. Several classes of DNA-binding factors display differential affinity for their target sequences depending on the dC modification state[1]. This is why genomic patterns of dC modification contribute to consolidation and propagation of gene expression states and chromatin structure, thus affecting fundamental processes such as cell fate determination, imprinting, X-chromosome inactivation and genome stability[1].

In the last two decades, methylation of genomic dC was shown to be rapidly removed in various biological contexts[6]. Several erasure mechanisms were proposed, including prevention of methylation maintenance, which results in passive dilution of the modified base through DNA replication, and active demethylation by enzymatic processes[6] (Fig. 1a). Among the latter, oxidation of mdC to fdC or cadC by Tet enzymes, followed by base excision repair (BER), mediated by the DNA glycosylases thymine DNA glycosylase (TDG)[5,7] and NEIL1 and NEIL2 (refs. [8,9]), is best characterized. Additional proposals invoke the involvement of other DNA repair pathways[6] or direct C–C bond cleavage[10–14]. A third line of proposal posits deamination of genomic mdC and hmdC to 2′-deoxythymidine (dT) and 5-hydroxymethyl-2′-deoxyuridine (hmdU), respectively, either by a cytosine deaminase of the AID/APOBEC family[15–17] or by de novo DNA methyltransferases DNMT3a and DNMT3b (ref. [18]). The resulting T:G or 5-hydroxymethyluracil (hmU):G mismatches would be then resolved by BER or non-canonical mismatch repair (ncMMR)[15,16,19,20]. Finally, it has been suggested that unmodified dC could be deaminated by AID, and long-patch BER or ncMMR could lead to indirect co-removal of adjacent mdC residues[17,19,20]. Currently, all of these pathways are highly controversial. Conclusive evidence for their functional relevance in vivo is lacking[6]. It is, however, conceivable that different pathways operate at different stages during development.

During early embryonic development, the epiblast transits through a spectrum of pluripotent states[21]. This transition can be modeled by cultivating mouse pluripotent stem cells (mPSCs) under specific conditions, which recapitulates dynamic expression and epigenetic changes, including a major global gain of mdC[21]. Conversely, reversion from partially and heterogeneously primed mPSC cultures to the naïve state involves downregulation of de novo methyltransferases and collapse of DNA methylation maintenance with consequent global and passive (independent of Tet enzymes) loss of genomic methylation[22]. In contrast, little is known about the extent and nature of the turnover of modified genomic cytosines during the forward, developmentally relevant transition from naïve to primed pluripotent states. Here, we used metabolic labeling with stable isotopes for highly accurate MS-based quantification of the global genomic mdC, hmdC and fdC turnover through a progression of pluripotent states. We show that, upon acquisition of primed pluripotency, roughly 3–6% of the methylome is actively turned over. Most of this active turnover is due to Tet-mediated mdC oxidation, while a minority involves an unknown mdC oxidation-independent DNA repair process.

Results

Genomic mdC oxidation kinetics decrease in primed mPSCs. We cultured mPSCs under specific conditions to model the progression of the embryonic epiblast from naïve to primed pluripotency

[1]Department of Chemistry, Ludwig Maximilians University Munich and Center for Integrated Protein Science Munich (CIPSM), Munich, Germany. [2]State Key Laboratory of Elemento-organic Chemistry and Department of Chemical Biology, College of Chemistry, Nankai University, Tianjin, China. [3]Present address: National Cancer Institute, Center for Cancer Research, Bethesda, MD, USA. [4]Present address: Cancer Research UK Cambridge Institute, Cambridge, UK. [5]Present address: Department of Chemistry, University of California, Los Angeles, Berkeley, CA, USA. ✉e-mail: fabio.spada@cup.lmu.de; thomas.carell@cup.lmu.de

and investigate the kinetics and turnover of dC modification during this transition. To monitor priming conditions, we first made use of an mPSC line expressing an Oct4-yellow fluorescent protein (YFP) reporter[23]. After adaptation to naïve conditions in the presence of Gsk3 inhibitor CHIR99021 (CHIR), Mek1 and Mek2 inhibitor PD0325901 (also known as 2i) and leukemia inhibitory factor (LIF), the cells were plated at low density in serum-containing medium supplemented with CHIR and the tankyrase inhibitor IWR-1-endo (serum/CR) as previously described[24]. Live-cell microscopy over the course of 6 d without passaging showed that these cells rapidly formed monolayer colonies, which retained a remarkably homogeneous nuclear YFP signal across the culture (Extended Data Fig. 1a), indicating that under these conditions mPSC cultures may not lose pluripotency. Similarly, intense and homogeneous YFP signals were observed when the same reporter line was permanently primed to mouse epiblast stem cells (mEpiSCs) in serum-free, chemically defined medium (CDM) supplemented with CHIR, IWR-1, fibroblast growth factor 2 (FGF-2) and activin A (CDM/CRFA; Extended Data Fig. 1a). Expression analysis of pluripotency factors in serum/CR cultures and CDM/CRFA mEpiSCs revealed similarly reduced and increased transcript levels of naïve and primed pluripotency factors, respectively (Extended Data Fig. 1b), showing that these conditions support similarly primed states. Thus, even without prolonged passaging or in chemically defined conditions, CHIR/IWR-1 treatment supports primed pluripotency.

We then used ultra-high pressure liquid chromatography coupled to triple quadrupole tandem mass spectrometry (UHPLC–MS^2) (ref. [25]) to analyze global levels of genomic mdC and hmdC upon priming of mPSCs under the same conditions over 5 d. As previously reported[26], the transition from the naïve state (2i/LIF) to a homogeneously primed pluripotent state involved a large increase in the global mdC levels, while the hmdC abundance dropped sharply, both modifications reaching plateau levels by day 4 (Extended Data Fig. 2a). Global levels of genomic mdC and hmdC, similar to those in serum/CR cultures, were detected in Oct4-YFP mEpiSCs under serum-free conditions, excluding potential confounding effects of undefined serum composition. Note that, under both serum-containing and serum-free conditions, we avoided supplementation with ascorbic acid, which greatly increases the levels of oxidized mdC derivatives[27]. The decrease of hmdC upon priming is consistent with a recent proteomic analysis showing that the levels of chromatin-bound Tet1 and Tet2 proteins are markedly reduced in mEpiSCs compared to naïve mPSCs, while Tet3 is undetectable in both states[28].

Notably, replacing the Mek inhibitor with the Src inhibitor CGP77675 (conditions referred to as alternative 2i or a2i) results in relatively high genomic mdC levels, while gene expression patterns remain similar to those of naïve mPSCs[29,30]. Our analysis showed that the expression pattern of pluripotency and early lineage specification factors under a2i/LIF conditions is intermediate between that of naïve and primed states. We found high levels of Tfcp2l1 and Klf4 transcripts relative to 2i/LIF (naïve) conditions, intermediate expression of Rex1, Stella and Prdm14, and low levels of Fgf5, Oct6, Otx2 and DNMT3b relative to primed conditions (Extended Data Fig. 1b). This pattern is similar to that present in formative pluripotency, an intermediate state along the trajectory from naïve to primed pluripotency, during which competence for multi-lineage commitment is acquired[31]. High-depth bisulfite amplicon sequencing revealed a gradual increase of CpG methylation at secondary imprints during the transition from a2i/LIF to serum/CR

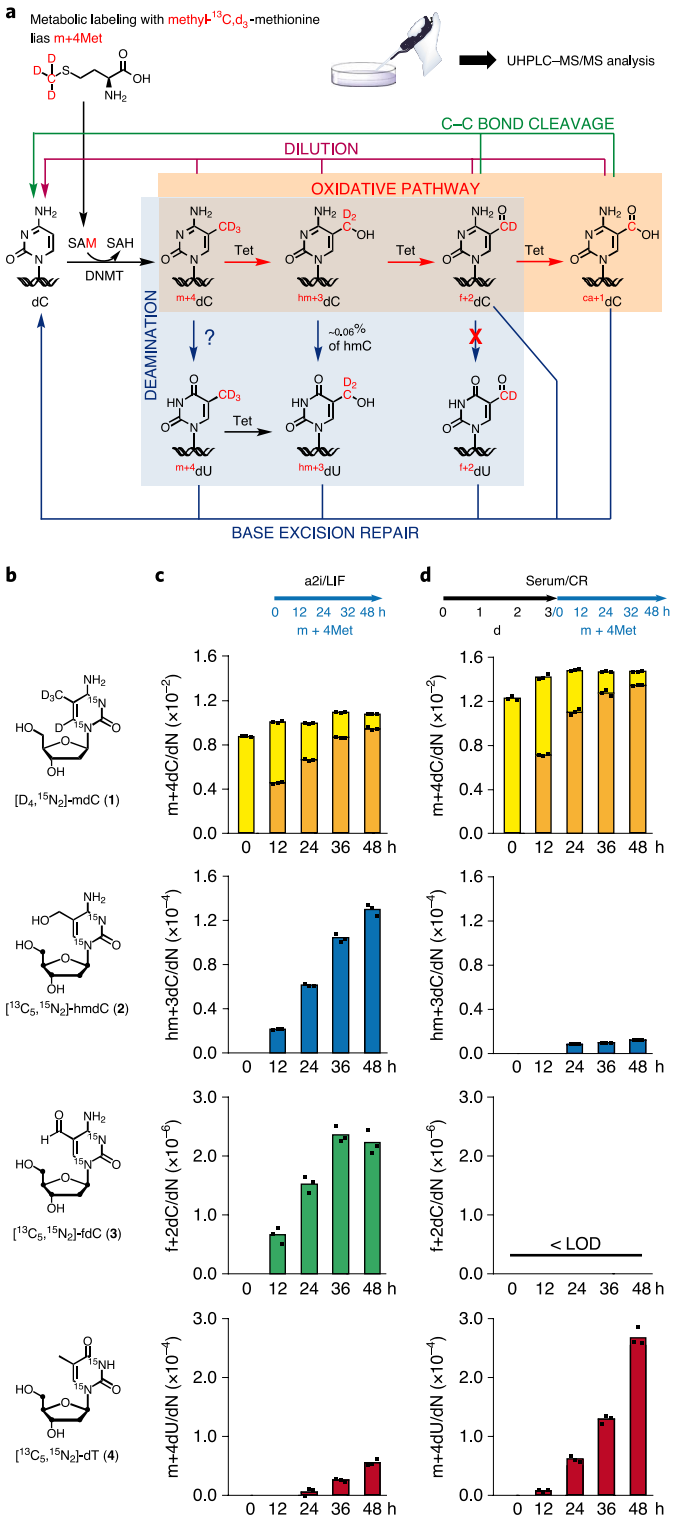


Fig. 1 | Drastic reduction of global genomic mdC oxidation and increase of mdC to mdU conversion upon transition to primed pluripotency. a, Proposed demodification pathways for genomic C5-cytosine modifications and expected labeled mdC derivatives upon metabolic labeling with m+4Met. **b**, Nucleoside isotopologues used as internal standards for UHPLC–MS^2 analysis. Black dots in the ribose represent ^{13}C atoms. **c**,**d**, Time-course analysis of genomic dC derivatives upon metabolic labeling with m+4Met under a2i/LIF (**c**) and priming conditions (serum/CR; **d**). The labeling/time-course schedules are shown at the top. Global levels of unlabeled mdC (pale yellow) and m+4dC (dark yellow), and hm+3dC (blue) and f+2dC (green) are shown as individual values of replicate measurements ($n = 3$; black squares) and their means (bars) from single samples. An independent biological replicate for **c** and **d** is shown in Extended Data Fig. 3.

conditions (Extended Data Fig. 2b), as observed in early embryos[32]. Altogether, these data support that cultures transferred from a2i/LIF to serum/CR conditions recapitulate fundamental molecular processes of the progression to a genuine primed pluripotent state. Importantly, a2i/LIF and serum/CR conditions allow comparison of genomic cytosine modification dynamics between functionally distinct pluripotent states with substantial and comparable steady-state levels of genomic mdC. Thus, we adopted a2i/LIF conditions both for routine mPSC maintenance and as a starting state for further time-course analyses of mPSC priming.

We next monitored global cytosine modification kinetics during acquisition of primed pluripotency by metabolic labeling of the C5 substituent. To this end, we supplemented L-methionine-free medium with isotopically labeled L-methionine (m+4Met), whereby the methyl group provides a +4 mass shift (methyl-$^{13}C,d_3$). As for naturally occurring L-methionine, m+4Met was incorporated into S-adenosylmethionine (SAM) and the $^{13}C,d_3$ (m+4) was then transferred to genomic dC by DNMTs, generating m+4dC. The latter can be sequentially oxidized by Tet proteins to hm+3dC, f+2dC and ca+1dC (Fig. 1a)[25].

For accurate quantification of natural mdC, hmdC and fdC and heavy m+4dC, hm+3dC and f+2dC, we synthesized a corresponding series of even heavier isotopologues (Fig. 1b; compounds 1–3) and used these as internal standards for quantification by UHPLC–MS^2. In this work, cadC and ca+1dC were not quantified because a difference of one mass unit cannot be discriminated. We first compared the modification kinetics in cultures maintained constantly under a2i/LIF conditions to cultures undergoing priming in CHIR/IWR-1 medium. In both cases, we performed time-course analyses over 2 d of labeling in the presence of m+4Met, which, in the case of priming cultures, was from days 3 to 5 after transfer to serum/CR medium (Fig. 1c,d and Extended Data Fig. 3). In priming cultures, the rate of m+4dC oxidation to hm+3dC was clearly slower than under a2i/LIF conditions, and accumulation of the further oxidized state f+2dC became undetectable upon priming. Thus, the transition to primed pluripotency determines a major decrease in global oxidation of genomic mdC, which is consistent with the drastically reduced steady-state levels of genomic hmdC (Extended Data Fig. 2a).

Oxidative turnover of genomic mdC in mPSCs. To monitor the turnover of modified genomic cytosines, we reversed the labeling schema described above. We first labeled cultures with m+4Met for 5 d, and, after transferring the cultures to medium with natural L-methionine, we measured the disappearance of the label over 2 d (Fig. 2a). Again, we compared cultures maintained constantly under serum/a2i/LIF conditions to cultures undergoing priming in serum/CR medium. For primed cultures, the first 2 pulse days were in a2i/LIF medium, while the last 3 d were in serum/CR medium, thus corresponding to the first 3 d of priming. To generate a reference for passive dilution of a genomic nucleoside through DNA synthesis, we co-labeled with an isotopically labeled dT having a mass shift of 12 units ($[^{15}N_2,^{13}C_{10}]$-dT; hereafter referred to as dT+12). Due to its highly efficient incorporation, dT+12 was co-pulsed only during the last day of the m+4Met pulse (Fig. 2a). Thereafter, dT+12 was chased together with m+4Met. Remarkably, while in most cases the levels of modified cytosines readily declined with the onset of the chase, under a2i/LIF conditions hm+3dC and f+2dC accumulated for the first 12 h and 24 h, respectively (Fig. 2b). This shows again that the rate of Tet-mediated oxidation is quite high in mPSCs under a2i/LIF conditions and declines sharply upon priming, as already seen with the time-course analysis during m+4Met labeling (Fig. 1b,c). Notably, under a2i/LIF conditions, the decline of f+2dC after 36 h was steep, pointing to high turnover of fdC. Although under priming conditions, oxidation of mdC to hmdC and fdC was hardly detectable in labeling time-course experiments (Fig. 1), in pulse-chase experiments, the decline of labeled hmdC and fdC was slower than that of dT+12, clearly showing that some mdC and hmdC oxidation still takes place (Fig. 2b). Importantly, upon priming, m+4dC turnover was faster than the dilution rate of dT+12 (Fig. 2c and Extended Data Fig. 4a), revealing the presence of a component that cannot be explained by passive dilution through DNA synthesis.

To assess the contribution of Tet-dependent oxidation to this active component, we performed the same pulse-chase experiment under priming conditions (Fig. 2a) with E14tg2a mPSCs lacking all three Tet proteins (Tet triple knockout; TKO)[33], as well as with parental Tet1–3 proficient E14tg2a cells (Fig. 3). In the parental cell line, turnover of genomic m+4dC was again clearly faster than passive dilution of dT+12. After averaging the data from three such pulse-chase experiments with wild-type mPSC lines, a paired two-tailed *t*-test showed high significance ($P=0.0053$) for the deviation between m+4dC and dT+12 decay rates. This accounted for an active component of global mdC turnover involving 3–6% of the methylome, depending on the time point (Fig. 2f). In contrast, in the genome of Tet TKO cells, the decline rate of m+4dC was essentially indistinguishable from that of dT+12 (Fig. 3). This shows that the active component of the mdC turnover is indeed largely dependent on the presence of Tet enzymes and thus mostly due to mdC oxidation.

Additional non-oxidative mdC turnover in mPSCs. Next to oxidative mdC turnover, we investigated potential deamination of mdC to dT (mdU)[15]. Upon metabolic labeling with m+4Met, deamination of m+4dC would generate m+4dU, that is, dT bearing the m+4 methyl group (Fig. 1a). To investigate this process quantitatively, we synthesized the isotopologue dT-$[^{13}C_5,^{15}N_2]$ (Fig. 1b; compound 4) and used it as a standard for UHPLC–MS^2 analysis. As the source of the methyl group for de novo dT biosynthesis is serine through 5,10-methylenetetrahydrofolate and not methionine through SAM, it was expected that, upon supplementation of the medium with m+4Met, m+4dC would be the only source of m+4dU. To verify this expectation, we made use of mPSCs lacking all catalytically active Dnmts (Dnmt TKO) and therefore devoid of genomic mdC[34]. After priming in the presence of m+4Met, m+4dU was clearly detected in the genome of wild-type mPSCs but not in Dnmt TKO cultures (Extended Data Fig. 3). Thus, generation of m+4dU strictly required the formation of genomic m+4dC, showing that m+4dU originates exclusively through deamination of m+4dC.

Direct deamination of genomic mdC would generate T:G mismatches, which, if not efficiently repaired, would be highly mutagenic. Despite this, we detected substantial accumulation of m+4dU in genomic DNA (gDNA), whenever cultures were provided with m+4Met (Figs. 1–5 and Extended Data Figs. 3–5 and 7). To explain this observation, we considered an alternative scenario in which m+4dC is released from the genome as soluble m+4dC/m+4dCMP, possibly through a DNA repair process. Deamination of soluble m+4dC/m+4dCMP by cytidine and deoxycytidylate deaminases (Cda and Dctd, respectively)[35], would generate soluble m+4dU/m+4dUMP, which could be phosphorylated and reincorporated into the genome through DNA synthesis (Fig. 4a). To test whether m+4dU was generated along this pathway, we supplemented the medium with m+4Met and increasing concentrations of unlabeled dT. The latter was added to outcompete the incorporation of soluble m+4dU/m+4dUMP into gDNA and to allosterically inhibit Dctd[36]. Indeed, we detected a dT dose-dependent reduction of genomic m+4dU in priming mPSC cultures (Fig. 4b). This result supports the idea that genomic m+4dU originates in large part from a pool of soluble precursors.

To probe whether deamination of m+4dC took place within gDNA or at the level of soluble nucleoside/nucleotide pools, we generated mPSCs with double deficiency for Cda and Dctd through CRISPR–Cas9-mediated gene editing (hereafter referred to as

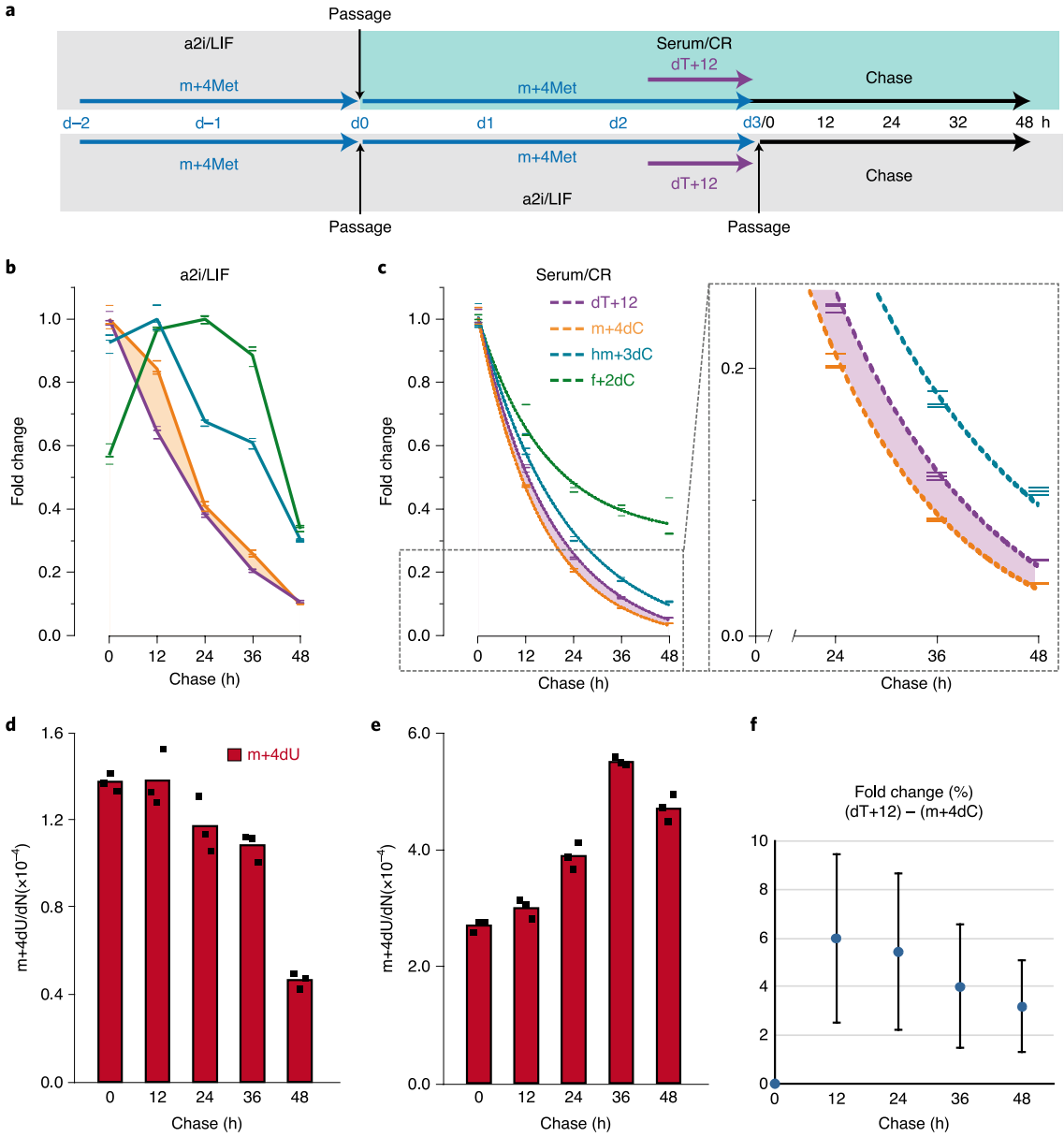


Fig. 2 | Turnover of genomic mdC and its derivatives under a2i/LIF and priming conditions. a, Workflow of m+4Met and dT+12 pulse-chase experiments with wild-type mPSCs under a2i/LIF and priming (serum/CR) conditions. **b**,**c**, Global profiles of the indicated labeled cytosine derivatives and dT+12 in the genome of mPSCs under a2i/LIF (**b**) and priming conditions (serum/CR; **c**) upon chasing of m+4Met. Each panel shows fold change values relative to the highest level for each modification from an individual experiment. Dashes represent single technical replicate measurements ($n = 3$) from single samples. Full lines in **b** join mean values, while dashed lines in **c** represent first-order decay curves fitted to mean values. In **b** and **c**, the area between the m+4dC and dT+12 lines is shadowed in yellow and purple, respectively. **d**,**e**, Absolute levels of genomic m+4dU from the same samples as in **b** and **c**, respectively. Black squares represent single technical replicate measurements ($n = 3$) from single samples, and bars show their mean values. The experiment under priming conditions (**c** and **e**) was independently repeated once in the same mPSC line (Extended Data Fig. 4) and once in E14tg2a mPSCs (Fig. 3). **f**, Percentage difference between fold changes of dT+12 and m+4dC in wild-type mPSCs under serum/CR conditions from independent biological experiments ($n = 3$; Figs. 2c and 3a and Extended Data Fig. 4a). Blue dots and error bars represent mean and s.d. values, respectively.

CD-DKO; Supplementary Fig. 1a,b). Genomic sequencing showed that three of the isolated clonal lines beared compound biallelic *Cda* and *Dctd* mutations incompatible with the expression of functional enzymes (Supplementary Figs. 1c–e and 2a,b). All three CD-DKO clones grew at slower rates than parental wild-type cells, a phenotype which was at least partially relieved by supplementation of dT (Supplementary Fig. 1f). This indicates that deamination of (d)C/dCMP by Cda and/or Dctd provides substantial amounts of (d)U/dUMP for biosynthesis of (d)T/dTMP by thymidine synthase to support DNA replication. This observation is fully consistent with previously reported dT auxotrophy of Dctd-deficient somatic cell lines[37]. As two of the CD-DKO clones grew particularly poorly under serum/CR priming conditions, we first used serum/LIF medium for metabolic labeling with m+4Met under priming conditions. Importantly, while the genomic content of mdC/m+4dC was the same as in parental wild-type cells, only background levels of m+4dU could be measured in all three CD-DKO clones (Fig. 4c). The same result was obtained with CD-DKO clone number 40 upon

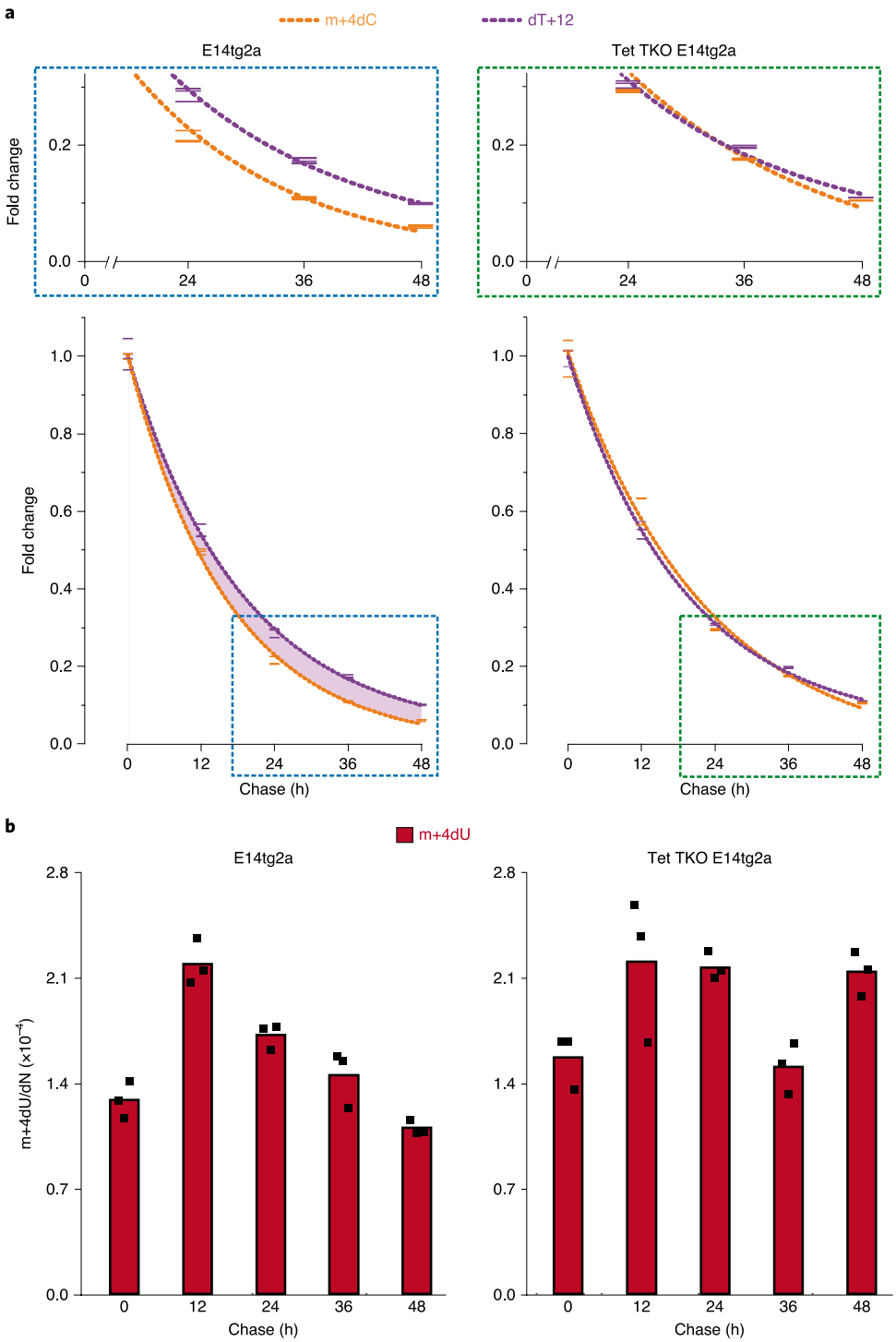


Fig. 3 | The active component of mdC turnover in primed pluripotent cells largely depends on Tet proteins. m+4Met and dT+12 pulse-chase experiment with parental and Tet TKO E14tg2a mPSCs under priming conditions (serum/CR; equivalent to Fig. 2c,e). Profiles of global genomic levels of the indicated labeled cytosine derivatives and dT+12 are shown. **a**, Fold change values for m+4dC and dT+12 relative to the highest data point for each modification. Upper panels display magnifications of dashed frames in lower panels. Dashes represent single technical replicate measurements ($n = 3$) from single samples, and dashed lines are first-order decay curves fitted to their means. In the middle panels, the area between the m+4dC and dT+12 lines is shadowed in purple. **b**, Absolute levels of genomic m+4dU from the same samples as in **a**. Black squares represent individual technical replicate measurements ($n = 3$) from single samples, and bars show their mean values. This experiment was not independently replicated.

priming in serum/CHIR/IWR-1 medium (Fig. 4d). These results clearly show that all measurable genomic m+4dU originates from a process involving direct release of m+4dC/m+4dCMP from the genome, its deamination to m+4dU/m+4dUMP in the soluble pool and reincorporation into the genome through phosphorylation and DNA synthesis.

Although our data do not support deamination of mdC directly within the genome, it is important to consider the limitations of our

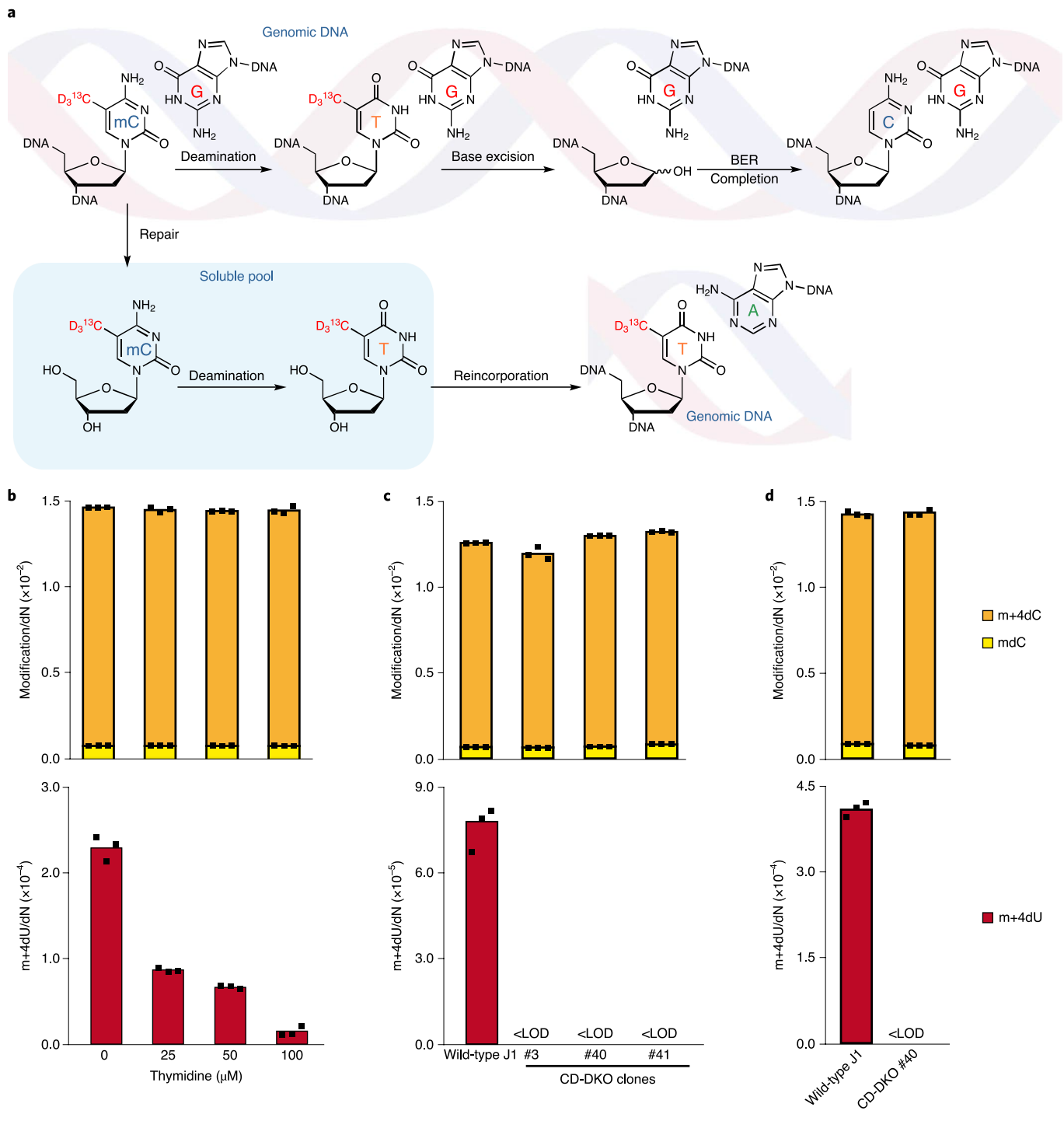


Fig. 4 | All detectable mdC deamination takes place in the soluble pool. a, Potential pathways for the conversion of genomic m+4dC into genomic m+4dU. **b**, Wild-type mPSCs were primed under CHIR/IWR-1 (serum/CR) conditions in the presence of m+4Met and growing concentrations of unlabeled dT in the culture medium. **c**, Wild-type and CD-DKO clones of mPSCs were grown under serum/LIF conditions in the presence of m+4Met. **d**, Wild-type and CD-DKO clone number 40 mPSCs were metabolically labeled with m+4Met under serum/CR priming conditions. Global genomic contents of mdC and m+4dC (upper) and m+4dU (lower) are shown as individual values of replicate measurements ($n = 3$: black squares) from single samples and their means (bars). In **b**–**d**, priming and labeling was for 5 d. Experiments in **b** and **c** were repeated three times with similar results.

UHPLC–MS^2 method. In UHPLC, the overwhelming amount of natural dT present in the genome co-elutes with m+4dU and outcompetes it during ionization for the capture of charge. This strong ion suppression effect reduces the m+4dU signal and results in a high limit of detection (LOD). To put this into perspective, our LOD for m+4dU (1×10^{-5} per nucleotide) is about 10–100 times higher than typical global levels of fdC, which is a relatively stable dC modification with an epigenetically relevant role[38,39]. In regard to this and previous reports of direct mdC deamination within gDNA[15,18], we cannot fully exclude that deamination of a few m+4dC residues within the genome triggers long-patch BER and/or ncMMR[19,20], leading to the release of larger amounts of m+4dC.

To minimize ion suppression and gain sensitivity, we double-labeled wild-type and CD-DKO mPSCs with m+4Met and a dC nucleoside where nine hydrogen atoms are replaced with deuterium. After incorporation into the genome, dC[D_9] is methylated to m+4dC[D_8], which deaminates to m+4dU[D_8] (a dT with eleven D atoms and a+12 mass shift; Extended Data Fig. 6a). This massive H-to-D exchange was expected to shift the UHPLC retention time slightly, allowing m+4dU[D_8] to escape co-elution with dT. Indeed, co-injection tests showed a slightly shorter retention time for dT[D_8] relative to dT (Extended Data Fig. 6b). For m+4dU[D_8] we saw an even shorter retention time. After double labeling with m+4Met and dC[D_9], we detected m+4dU[D_8] confidently in wild-type mPSCs (Extended Data Fig. 6c). In addition, we sporadically detected small m+4dU[D_8] peaks in technical replicates of CD-DKO samples. After integration of these sporadic signals, the corresponding values were invariably below the LOD. Although we cannot exclude that these signals were caused by background processes, taking them into consideration allowed us to estimate between 500 and 5,000 genomic deamination events per cell in the course of our labeling experiment. According to an empirical rate constant for spontaneous hydrolytic deamination of mdC in solution[40], only around 20 such events would be expected to take place. Therefore, we cannot exclude the occurrence of a small amount of genomic deamination above the rate of spontaneous hydrolysis at levels of 1×10^{-6} to 1×10^{-7} per nucleotide.

Tet enzymes do not trigger mdC-to-dT turnover. Conversion of m+4dC into m+4dU does not require oxidized intermediates. However, it was recently suggested that Tet3 generates undefined DNA lesions in the zygote, triggering repair-mediated erasure of genomic methylation from the paternal genome[41]. Also, hmdC may mark DNA damage sites and promote their repair[42] or even elicit DNA repair by selectively recruiting endonucleases such as Srap1 (ref. [43]) and endonuclease G[44]. Therefore, we considered whether Tet proteins contributed to the turnover of genomic mdC into dT by generating hmdC or low levels of DNA lesions such as hmdU[25], which triggers BER initiated by the DNA glycosylase Smug1. To this aim, we primed Tet TKO and parental E14tg2a mouse embryonic stem cells in the presence of m+4Met. Surprisingly, higher levels of genomic m+4dU were detected in the absence of Tet enzymes (Extended Data Fig. 7). Similarly, analysis of the pulse-chase experiment also revealed a tendency for higher accumulation of m+4dU in Tet TKO mPSCs at later chase time points (Fig. 3b). Therefore, in pluripotent cells, Tet proteins do not seem to trigger DNA repair events that may contribute to turnover of mdC into dT.

mdC-to-dT turnover is developmentally regulated. To investigate whether oxidation-independent mdC turnover is regulated during the transition to primed pluripotency, we first compared the genomic abundance of m+4dU in the time-course and pulse-chase experiments under the a2i/LIF and priming conditions described above (Figs. 1–3 and Extended Data Fig. 3). In time-course experiments, progressive accumulation of m+4dU was much faster and about five times higher under priming conditions (Fig. 1b and Extended Data Fig. 3). Also, m+4dU was the only m+4dC derivative that accumulated during label chase under priming conditions, although to different extents in the two cell lines tested (Figs. 2e and 3b and Extended Data Fig. 4b). Thus, in contrast to the observed decline of mdC oxidation (Figs. 1 and 2 and Extended Data Fig. 3), these results reveal increased rates of mdC-to-dT turnover upon acquisition of primed pluripotency. This anti-correlation

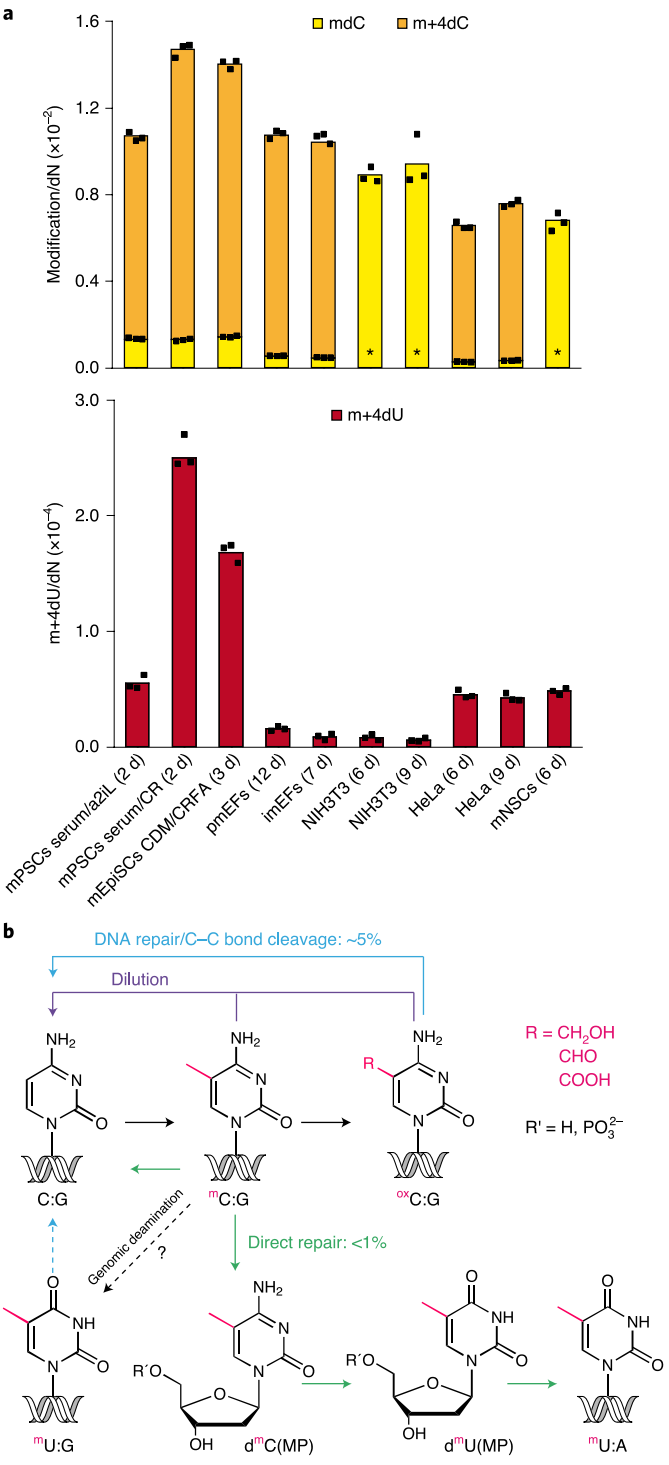


Fig. 5 | Oxidation-independent mdC-to-dT turnover is developmentally regulated. a, Cultures of mPSCs in a2i/LIF and primed states, pmEFs, imEFs, NIH3T3, HeLa cells and mPSC-derived NSCs (mNSCs) were labeled for the indicated number of days with m+4Met. The asterisks indicate cases where distinct parallel cultures in the absence and presence of m+4Met were used to measure global levels of genomic mdC and m+4dU, respectively. In all other cases, mdC, m+4dC and m+4dU were measured from the same culture labeled with m+4Met. Individual values of replicate measurements ($n=3$: black squares) from single samples and their means (bars) are shown. Only the first two samples were independently replicated a second time. a2iL, a2i/LIF; CR, CHIR/IWR-1; FA, FGF-2/activin A. **b**, Schematic summary of the dC modification turnover pathways identified in primed pluripotent cells. Deamination of genomic mdC could not be detected and may occur only at levels below the sensitivity of our assay.

suggests that oxidation-independent mdC turnover may serve, at least in part, as a supplementary pathway for the erasure of dC methylation. However, this process cannot involve more than 1% of the methylome, because in primed Tet TKO mPSCs there was no apparent difference between global mdC turnover and DNA replication-dependent passive dilution (Fig. 3a; decay of m+4dC and dT+12, respectively).

At last, we compared oxidation-independent mdC-to-dT turnover in pluripotent states with various lineage-committed cell types, including primary and immortalized embryonic fibroblasts (pmEFs and imEFs, respectively), neural stem cells (NSCs)[45] and HeLa cells. Although primed pluripotent cells were labeled with m+4Met for only 2 to 3 d, m+4dU reached higher levels in their genome than in those of any of the somatic cell types labeled for substantially longer periods, ranging from 6 to 12 d (Fig. 5a). This clearly shows that, in lineage-committed cells, oxidation-independent mdC turnover is substantially lower than in primed pluripotent cells. In addition, the rate of mdC-to-dT turnover is clearly not proportional to the global genomic abundance of mdC either between pluripotent states or among pluripotent and lineage-committed cells. In particular, among the latter, mdC-to-dT turnover is clearly more active in NSCs than in fibroblasts, despite the mdC genomic content being at least 20% lower in NSCs. These observations indicate that mdC-to-dT turnover is under developmental and cell lineage-dependent control.

Discussion

Recent methylome analysis of matched clonal and polyclonal cell populations[46] and studies on the contribution of DNMT3 proteins to DNA methylation maintenance[47,48] suggest that, in heterogeneously primed mPSC cultures (serum/LIF), genomic mdC undergoes higher dynamic turnover as compared to differentiated somatic cells. However, the extent and especially the nature of this turnover have not been addressed. Previously, Bachmann et al. reported a pulse-chase experiment with m+4Met and a short pulse time to estimate the global turnover of genomic mdC and hmdC in mPSCs under naïve conditions (2i/LIF)[49], where the levels of dC modifications are far lower.

Here, we report the analysis of on- and off-rate kinetics for genomic dC modifications upon transition to primed pluripotency. We show that, in mPSCs under a2i/LIF conditions, which display expression and genomic methylation patterns that are intermediate between naïve and primed pluripotent states (formative state)[31], the rates of mdC oxidation to hmdC and further to fdC are quite sustained. The removal/conversion of fdC is very rapid. This points to a dynamic equilibrium for fdC under these conditions. It is tempting to speculate that these high levels of oxidative turnover may underlie the frequent periodical oscillations in global mdC levels recently identified through single-cell analysis in partially (and heterogeneously) primed serum/LIF cultures[50]. In contrast to the high levels of mdC oxidation in a2i/LIF conditions, upon transition to primed pluripotency, the global oxidation rates of mdC and hmdC, as well as the removal of hmdC and fdC decline drastically. Despite this, the global turnover rate of genomic mdC is 3–6% faster than can be accounted for by passive dilution through DNA synthesis (DNA replication and repair; Fig. 5b). This difference is largely due to Tet-mediated mdC oxidation. Evidence for passive erasure of dC methylation based on impairment of its maintenance has been provided for primordial germ cells, preimplantation embryonic development and the reversion from heterogeneous partially primed/formative states present in serum/LIF cultures to the naïve pluripotent state[1,22]. We suggest that, against the backdrop of global gain and maintenance of dC methylation in the forward developmental transition from naïve to primed pluripotency, an active demethylation mechanism may be more attainable than preventing maintenance of dC methylation at selected sites. This could be especially the case at loci with a highly accessible chromatin conformation like actively transcribed genes.

By tracing the origin of genomic m+4dU to the deamination of soluble m+4dC, we show that a small part of genomic mdC is turned over by a process that involves the direct excision of mdC from the genome (Fig. 5b). Although the nature of this process and the events triggering it remain to be determined, we speculate that DNA repair is responsible for the release mdC. While we show that Tet proteins are clearly not involved in triggering or targeting this non-oxidative mdC turnover, our data neither support nor rule out that low levels of direct enzymatic deamination of genomic mdC may trigger a DNA repair process underpinning the turnover pathway. Importantly, this process is not proportional to global genomic mdC levels between pluripotent states, among lineage-committed cell types or between pluripotent and committed cells, indicating that it could be developmentally regulated. Interestingly, the levels of oxidation-independent mdC turnover show an inverse correlation with mdC oxidation rates across pluripotent states and seem to be slightly more sustained in Tet-deficient cells, pointing to a possible function as an alternative or supplementary mechanism dedicated to active erasure of cytosine methylation.

In summary, we show that upon priming of pluripotent cells, a significant fraction of mdC (3–6%) vanishes after oxidation to hmdC, fdC and cadC. A smaller fraction of mdC is removed from the genome by cleavage of phosphodiester bonds, likely as part of an unknown DNA repair process. Once released in the soluble pool, mdC is deaminated to dT, which is reintegrated into the genome. Finally, our data are inconclusive regarding a very low level of genomic mdC deamination, which would generate dT:dG mismatches and trigger DNA repair.

Online content

Any Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41589-020-0621-y.

Received: 14 November 2019; Accepted: 8 July 2020;
Published online: 10 August 2020

References

1. Greenberg, M. V. C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* 20; 590–607 (2019).
2. He, Y. & Ecker, J. R. Non-CG methylation in the human genome. *Annu. Rev. Genomics Hum. Genet.* **16**, 55–77 (2015).
3. Tahiliani, M. et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner Tet1. *Science* **324**, 930–935 (2009).
4. Ito, S. et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
5. He, Y.-F. et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).
6. Schuermann, D., Weber, A. R. & Schär, P. Active DNA demethylation by DNA repair: facts and uncertainties. *DNA Repair* **44**, 92–102 (2016).
7. Maiti, A. & Drohat, A. C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.* **286**, 35334–35338 (2011).
8. Slyvka, A., Mierzejewska, K. & Bochtler, M. Nei-like 1 (NEIL1) excises 5-carboxylcytosine directly and stimulates TDG-mediated 5-formyl and 5-carboxylcytosine excision. *Sci. Rep.* **7**, 9001 (2017).
9. Schomacher, L. et al. NEIL DNA glycosylases promote substrate turnover by Tdg during DNA demethylation. *Nat. Struct. Mol. Biol.* **23**, 116–124 (2016).
10. Chen, C.-C., Wang, K.-Y. & Shen, C.-K. J. The mammalian de novo DNA methyltransferases Dnmt3a and Dnmt3b Are Also DNA 5-hydroxymethyl cytosine dehydroxymethylases. *J. Biol. Chem.* **287**, 33116–33121 (2012).
11. Liutkevičiūtė, Z., Lukinavičius, G., Masevičius, V., Daujotytė, D. & Klimašauskas, S. Cytosine-5-methyltransferases add aldehydes to DNA. *Nat. Chem. Biol.* **5**, 400–402 (2009).

12. Iwan, K. et al. 5-formylcytosine to cytosine conversion by C–C bond cleavage in vivo. *Nat. Chem. Biol.* **14**, 72–78 (2017).
13. Liutkevičiūtė, Z. et al. Direct decarboxylation of 5-carboxylcytosine by DNA C5-methyltransferases. *J. Am. Chem. Soc.* **136**, 5884–5887 (2014).
14. Schiesser, S. et al. Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. *Angew. Chem. Int. Ed. Engl.* **51**, 6516–6520 (2012).
15. Rai, K. et al. DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase and Gadd45. *Cell* **135**, 1201–1212 (2008).
16. Cortellino, S. et al. Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146**, 67–79 (2011).
17. Santos, F. et al. Active demethylation in mouse zygotes involves cytosine deamination and base excision repair. *Epigenetics Chromatin* **6**, 39 (2013).
18. Metivier, R. et al. Cyclical DNA methylation of a transcriptionally active promoter. *Nature* **452**, 45–50 (2008).
19. Grin, I. & Ishchenko, A. A. An interplay of the base excision repair and mismatch repair pathways in active DNA demethylation. *Nucleic Acids Res.* **44**, 3713–3727 (2016).
20. Franchini, D.-M. et al. Processive DNA demethylation via DNA deaminase-induced lesion resolution. *PLoS ONE* **9**, e97754 (2014).
21. Atlasi, Y. & Stunnenberg, H. G. The interplay of epigenetic marks during stem cell differentiation and development. *Nat. Rev. Genet.* **18**, 643–658 (2017).
22. von Meyenn, F. et al. Impairment of DNA methylation maintenance is the main cause of global demethylation in naive embryonic stem cells. *Mol. Cell* **62**, 848–861 (2016).
23. Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K. & Niwa, H. Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development* **135**, 909–918 (2008).
24. Kim, H. et al. Modulation of β-catenin function maintains mouse epiblast stem cell and human embryonic stem cell self-renewal. *Nat. Commun.* **4**, 2403 (2013).
25. Pfaffeneder, T. et al. Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat. Chem. Biol.* **10**, 574–581 (2014).
26. Shirane, K. et al. Global landscape and regulatory principles of DNA methylation reprogramming for germ cell specification by mouse pluripotent stem cells. *Developmental Cell* **39**, 87–103 (2016).
27. Blaschke, K. et al. Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* **500**, 222–226 (2013).
28. van Mierlo, G., Wester, R. A. & Marks, H. A mass spectrometry survey of chromatin-associated proteins in pluripotency and early lineage commitment. *Proteomics* **19**, e1900047 (2019).
29. Yagi, M. et al. Derivation of ground-state female ES cells maintaining gamete-derived DNA methylation. *Nature* **548**, 224–227 (2017).
30. Choi, J. et al. Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells. *Nature* **548**, 219–223 (2017).
31. Kalkan, T. et al. Tracking the embryonic stem cell transition from ground state pluripotency. *Development* **144**, 1221–1234 (2017).
32. Wang, L. et al. Programming and inheritance of parental DNA methylomes in mammals. *Cell* **157**, 979–991 (2014).
33. Lu, F., Liu, Y., Jiang, L., Yamaguchi, S. & Zhang, Y. Role of Tet proteins in enhancer activity and telomere elongation. *Genes Dev.* **28**, 2103–2119 (2014).
34. Tsumura, A. et al. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11**, 805–814 (2006).
35. Jekunen, A., Puukka, M. & Vilpo, J. Exclusion of exogenous 5-methyl-2'-deoxycytidine from DNA in human leukemic cells: A study with [2-14C]- and [methyl-14C]5-methyl-2'-deoxycytidine. *Biochem. Pharmacol.* **32**, 1165–1168 (1983).
36. Maley, G. F., Lobo, A. P. & Maley, F. Properties of an affinity-column-purified human deoxycytidylate deaminase. *Biochim. Biophys. Acta* **1162**, 161–170 (1993).
37. de Saint Vincent, B. R., Déchamps, M. & Buttin, G. The modulation of the thymidine triphosphate pool of Chinese hamster cells by dCMP deaminase and UDP reductase. Thymidine auxotrophy induced by CTP in dCMP deaminase-deficient lines. *J. Biol. Chem.* **255**, 162–167 (1980).
38. Bachman, M. et al. 5-formylcytosine can be a stable DNA modification in mammals. *Nat. Chem. Biol.* **11**, 555–557 (2015).
39. Bilyard, M. K., Becker, S. & Balasubramanian, S. Natural, modified DNA bases. *Curr. Opin. Chem. Biol.* **57**, 1–7 (2020).
40. Shen, J.-C., Rideout, W. M. & Jones, P. A. The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Res.* **22**, 972–976 (1994).
41. Ladstätter, S. & Tachibana-Konwalski, K. A surveillance mechanism ensures repair of DNA lesions during zygotic reprogramming. *Cell* **167**, 1774–1787 (2016).
42. Jiang, D., Wei, S., Chen, F., Zhang, Y. & Li, J. Tet3-mediated DNA oxidation promotes ATR-dependent DNA damage response. *EMBO Rep.* **18**, 781–796 (2017).
43. Kweon, S.-M. et al. Erasure of Tet-oxidized 5-methylcytosine by a SRAP nuclease. *Cell Rep.* **21**, 482–494 (2017).
44. Robertson, A. B., Robertson, J., Fusser, M. & Klungland, A. Endonuclease G preferentially cleaves 5-hydroxymethylcytosine-modified DNA creating a substrate for recombination. *Nucleic Acids Res.* **42**, 13280–13293 (2014).
45. Liu, N. et al. Intrinsic and extrinsic connections of Tet3 dioxygenase with CXXC zinc finger modules. *PLoS ONE* **8**, e62755 (2013).
46. Shipony, Z. et al. Dynamic and static maintenance of epigenetic memory in pluripotent and somatic cells. *Nature* **513**, 115–119 (2014).
47. Ooi, S. K. et al. Dynamic instability of genomic methylation patterns in pluripotent stem cells. *Epigenetics Chromatin* **3**, 17 (2010).
48. Chen, T., Ueda, Y., Dodge, J. E., Wang, Z. & Li, E. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol. Cell. Biol.* **23**, 5594–5605 (2003).
49. Bachman, M. et al. 5-hydroxymethylcytosine is a predominantly stable DNA modification. *Nat. Chem.* **6**, 1049–1055 (2014).
50. Rulands, S. et al. Genome-scale oscillations in DNA methylation during exit from pluripotency. *Cell Syst.* **7**, 63–76 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

Methods

Cell culture. Media components were from Sigma unless specified otherwise. Basal medium for mPSC culture was high glucose DMEM containing 10% FBS (Pan-ES, Pan-Biotech), 2 mM L-alanyl-L-glutamine, 1× MEM non-essential amino acid solution and 0.1 mM β-mercaptoethanol. mPSC lines were routinely maintained in basal medium supplemented with 1,000 U ml^{-1} LIF (ORF Genetics), 3 µM CHIR99021 and 7.5 µM CGP77675 (a2i). For adaptation to 2i/LIF conditions, CGP was replaced with 1 µM PD0325901. For priming, mPSC basal medium was supplemented with 1.5 µM CHIR99021 and IWR-1-endo at 2.5 µM as previously reported[24,51]. For experiments under serum/LIF conditions, basal medium was supplemented exclusively with 1,000 U ml^{-1} LIF. Small-molecule inhibitors were purchased from Selleckchem (CHIR, PD and IWR-1), Axon Medchem (CHIR, CGP, PD and IWR-1) or MedChemExpress (CHIR). Gelatin coating was used for all mPSCs cultures in serum-containing medium. Unless otherwise stated, experiments with wild-type mPSCs were performed with an embryonic stem cell line derived from 129 and C57BL/6 F1 hybrid embryos[52]. For metabolic labeling experiments with m+4Met under all conditions, mPSC basal medium was generated with L-methionine- and L-cysteine-free DMEM (Sigma or Life Technologies) supplemented with 0.2 mM m+4Met ([methyl-^{13}C,d$_3$]) and L-cysteine (both from Sigma). In pulse-chase experiments, dT+12 ([^{15}N$_2$,^{13}C$_{10}$]-dT; BACH UG) was supplemented at 5 µM. dC[D$_9$] (Silantes) was supplemented to culture medium at 100 µM. mEpiSCs in serum-free CDM were derived from the Oct4-YFP reporter mPSC line OLY2-1 (ref. [23]). These cells were first adapted to CDM supplemented with CHIR, PD and LIF as reported above, then transferred to CDM supplemented with CHIR, CGP, 12 ng ml^{-1} FGF-2 (Miltenyi Biotec) and 20 ng ml^{-1} activin A (PeproTec) and maintained in the latter medium by passaging as small cell clusters using 5 mM EDTA in Hank's balanced salt solution without Mg^{2+}, Ca^{2+} and sodium carbonate supplemented with 10 mM HEPES buffer. CDM was high glucose DMEM supplemented with 1× NEAA, 2 mM L-Alanyl-L-Glutamine, 15 µg ml^{-1} human recombinant insulin, 10 µg ml^{-1} human holo-transferrin (Merck), 12.5 mg ml^{-1} AlbuMAX I (Life Technologies), 0.7 µg ml^{-1} vitamin B$_{12}$, 1.8 ng ml^{-1} biotin, 0.75 µM $ZnSO_4$ and 2.6 nM $CuSO_4$. Laminin 521 coating (BioLamina) was used for all mPSC cultures in CDM. The neural stem cell line ENC1 was generated and cultured as previously described[45]. ENC1 cells were labeled with m+4Met ([methyl-^{13}C,d$_3$]) in DMEM/F-12 generated with L-Methionine-free DMEM supplemented with N2 (ref. [53]) and 20 ng ml^{-1} each of FGF-2 and EGF (PerpoTech). Primary mEFs (CF-1, Applied StemCell) were cultured in mPSC basal medium. Fibroblast cell lines and HeLa cells were cultured in DMEM containing 10% FBS (Life Technologies), 2 mM L-glutamine and 0.1 mM β-mercaptoethanol.

Generation of CD-DKO mPSCs by CRISPR–Cas9 editing. Single-guide RNAs (sgRNAs) were designed over exon1 and 4 of *Cda* and *Dctd*, respectively (Supplementary Fig. 1), using the CRISPR design tool (http://crispr.mit.edu/). Oligonucleotides used for construction of Cas9/sgRNA expression vectors are listed in Supplementary Table 1. The oligonucleotides were annealed and cloned in the BbsI site of pSpCas9-2A-Puro (PX459; Addgene Plasmid 48139)[54]. Cas9/sgRNA expression vectors targeting both *Cda* and *Dctd* were co-transfected in wild-type J1 mPSCs using Lipofectamine 2000 (Life Technologies) according to the manufacturer's instructions. Two days after transfection, cultures were selected with 1 µg ml^{-1} puromycin for 2 d and subcloned by limiting dilution. Clones were screened using the Surveyor mutation detection kit (Transgenomics), and potential compound biallelically targeted clones were subject to sequencing. For reverse transcription followed by PCR (RT–PCR), RNA was isolated with the ZR-Duet DNA/RNA MiniPrep kit (Zymo Research) and first-strand cDNA synthesis was performed as previously described[55]. Primers used to amplify the Cda cDNA are reported in Supplementary Table 1. For western blot analysis of Dctd, whole-cell extracts and purified human recombinant DCTD protein[56] were probed with a rabbit anti-DCTD antibody[36].

Isolation of gDNA and UHPLC–MS^2 analysis. Isolation, digestion and UHPLC–MS^2 analysis of gDNA samples were performed as described previously[25,57] with the following modifications: for every technical replicate 4 µg of gDNA was digested and analyzed; the enzyme mixture including S1 nuclease, Antarctic phosphatase and snake venom phosphodiesterase was used for digestion; the digestion mixture was supplemented with 0.8 µM tetrahydrouridine (Abcam) to inhibit potential cytidine deaminase activity in enzyme preparations.

RT–quantitative PCR analysis. Total RNA was isolated from the lysate flow-throughs obtained upon gDNA isolation with silica spin columns. RNA isolation from the flow-throughs was performed as described in the Quick-DNA/RNA MiniPrep kit (Zymo Research, D7001). In-column DNase I digestion was performed to minimize carryover of residual gDNA. One microgram of total RNA was reverse transcribed with the iScript cDNA synthesis kit and the cDNA product equivalent to 50 ng of total RNA was amplified using iTaq Universal SYBR Green Supermix (Bio-Rad) on a qTOWER^3/G cycler (Jena Biosciences). Quantitative PCR (qPCR) primers are reported in Supplementary Table 1. Each primer pair spanned across one intron except for Oct6 cDNA amplifications as *Pou3f1* (the gene coding Oct6) is intronless. qPCR reactions were also performed in parallel for every target/sample by omitting the reverse transcriptase to control for signals generated from residual gDNA (intronless Oct6 transcript, relatively short introns and spliced pseudogenes). An RT–qPCR assay for Gapdh transcripts was used as a housekeeping transcript reference to calculate ΔCt values. Fold change values were calculated with the ΔΔCt method.

High-depth targeted bisulfite amplicon sequencing and analysis. gDNA (1–2 µg) was treated with bisulfite using the EpiTect bisulfite kit (Qiagen, 59104). 5′ overhangs compatible with Illumina TruSeq and Nextera adaptors were appended to the locus-specific primers (Supplementary Table 1). Approximately 10 ng of bisulfite-converted DNA was amplified with HotStar-Taq Plus DNA Polymerase (Qiagen). PCR products were analyzed by agarose gel electrophoresis and quantified by densitometry using ImageJ (v1.52d). Similar amounts of these primary amplicons were used as templates for a second round of amplification with i5 and i7 indexing primers (Supplementary Table 1). PCR products were again analyzed by agarose gel electrophoresis and, after purification with CleanNGS magnetic beads (CleanNA), they were quantified by fluorometry with Quant-iT PicoGreen reagent (Thermo Fisher). Equimolar amounts of amplicons were then pooled, and the size distribution of the final library was assessed on a Bioanalyzer (Agilent). The 2×300 output mode for dual-indexed sequencing was used on a MiSeq instrument (Illumina). Size, numbers of CpG sites and genomic coordinates of targeted regions are reported in Supplementary Table 2. Sequenced reads were trimmed for adaptor sequences and masked for low-quality sequence using TrimGalore (v0.5.0) and Cutadapt (v1.17). Trimmed reads were mapped to imprint-containing, bisulfite-converted chromosomes 2, 7, 11, 12 and 17 from mm10 using bsmap (v2.90). Methylation ratios were extracted from bam files using a custom Python script and tabulated as a matrix table from which the box plot in Extended Data Fig. 2b was generated by means of an R script.

Synthetic procedures. The synthesis of labeled nucleosides standards is reported in the Supplementary Note.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The raw sequencing data, matrix table and targeted bisulfite amplicon sequencing data for Extended Data Fig. 2b are available at the Gene Expression Omnibus (GEO) under accession code GSE152174. The authors declare that all other data supporting the findings of this study are available within the paper and its Supplementary Information files.

References

51. Zhou, X., Chadarevian, J. P., Ruiz, B. & Ying, Q.-L. Cytoplasmic and nuclear TAZ exert distinct functions in regulating primed pluripotency. *Stem Cell Rep.* **9**, 732–741 (2017).
52. Moser, M., Nieswandt, B., Ussar, S., Pozgajova, M. & Fässler, R. Kindlin-3 is essential for integrin activation and platelet aggregation. *Nat. Med.* **14**, 325–330 (2008).
53. Nichols, J. & Ying, Q.-L. Derivation and propagation of embryonic stem cells in serum- and feeder-free culture. in *Embryonic Stem Cell Protocols* (ed. Turksen, K.) Vol. 329, 91–98 (Humana Press, 2006).
54. Ran, F. A. et al. Genome engineering using the CRISPR–Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
55. Rahimoff, R. et al. 5-formyl- and 5-carboxydeoxycytidines do not cause accumulation of harmful repair intermediates in stem cells. *J. Am. Chem. Soc.* **139**, 10359–10364 (2017).
56. Weiner, K. X., Weiner, R. S., Maley, F. & Maley, G. F. Primary structure of human deoxycytidylate deaminase and overexpression of its functional protein in *Escherichia coli*. *J. Biol. Chem.* **268**, 12983–12989 (1993).
57. Traube, F. R. et al. Isotope-dilution mass spectrometry for exact quantification of non-canonical DNA nucleosides. *Nat. Protoc.* **14**, 283–312 (2019).

Acknowledgements

We are very thankful to the following colleagues: M. Möser (Max Planck Institute for Biochemistry) for K3 mPSCs (kindlin3$^{+/+}$); H. Niwa and M. Okano (both at Kumamoto University) for Oct4-YFP-Puro E14tg2a and Dnmt TKO J1 mPSCs, respectively; Y. Zhang (Boston Children's Hospital) for parental and Tet TKO E14tg2a cells; and S. Bultmann and C. Mulholland (both at Ludwig Maximilian University) for guidance on high-depth bisulfite amplicon sequencing. Anti-DCTD antibody and purified recombinant DCTD protein were generous gifts from F. Maley (New York State University). Funding was provided by the Deutsche Forschungsgemeinschaft via the programs SFB1309 (TP: A4), SFB1361 (TP: 2), GRK 2338 and SPP1784.

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. EPiR 741912). Y.Z. is supported by the China Scholarship Council (CSC no. 201806200069).

Author contributions

F.S. and T.C. conceived the study and designed the experiments. F.S., S.S., A.K., Y.Z., G.A. and O.K. performed experiments and analyzed data. F.S., T.C. and S.S. interpreted data. R.R., C.E. and E.K. synthesized isotopically labeled nucleosides used as standards for LC–MS^2. F.S., T.C. and S.S. wrote the manuscript.

Competing interests

The authors declare no competing interests.

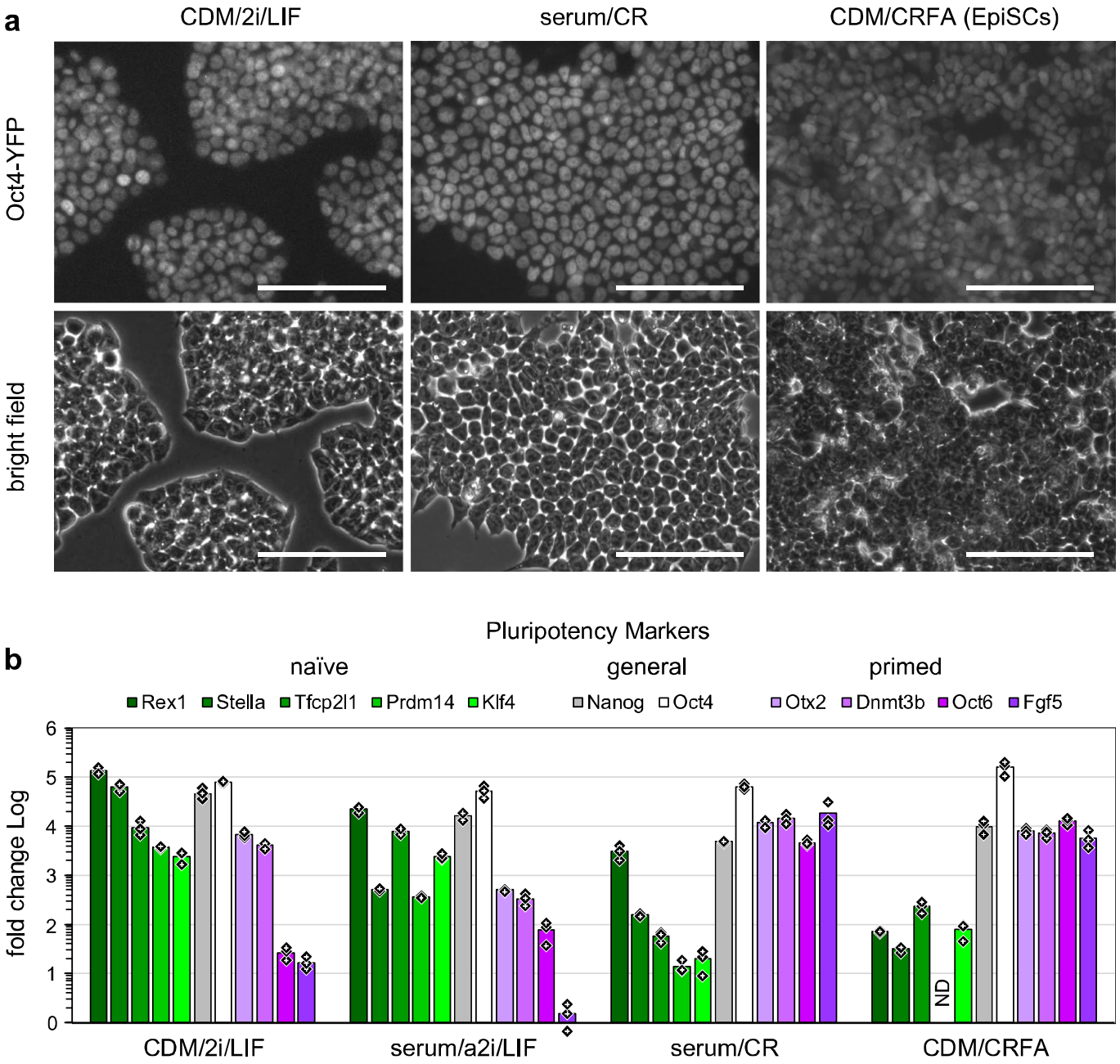
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41589-020-0621-y.

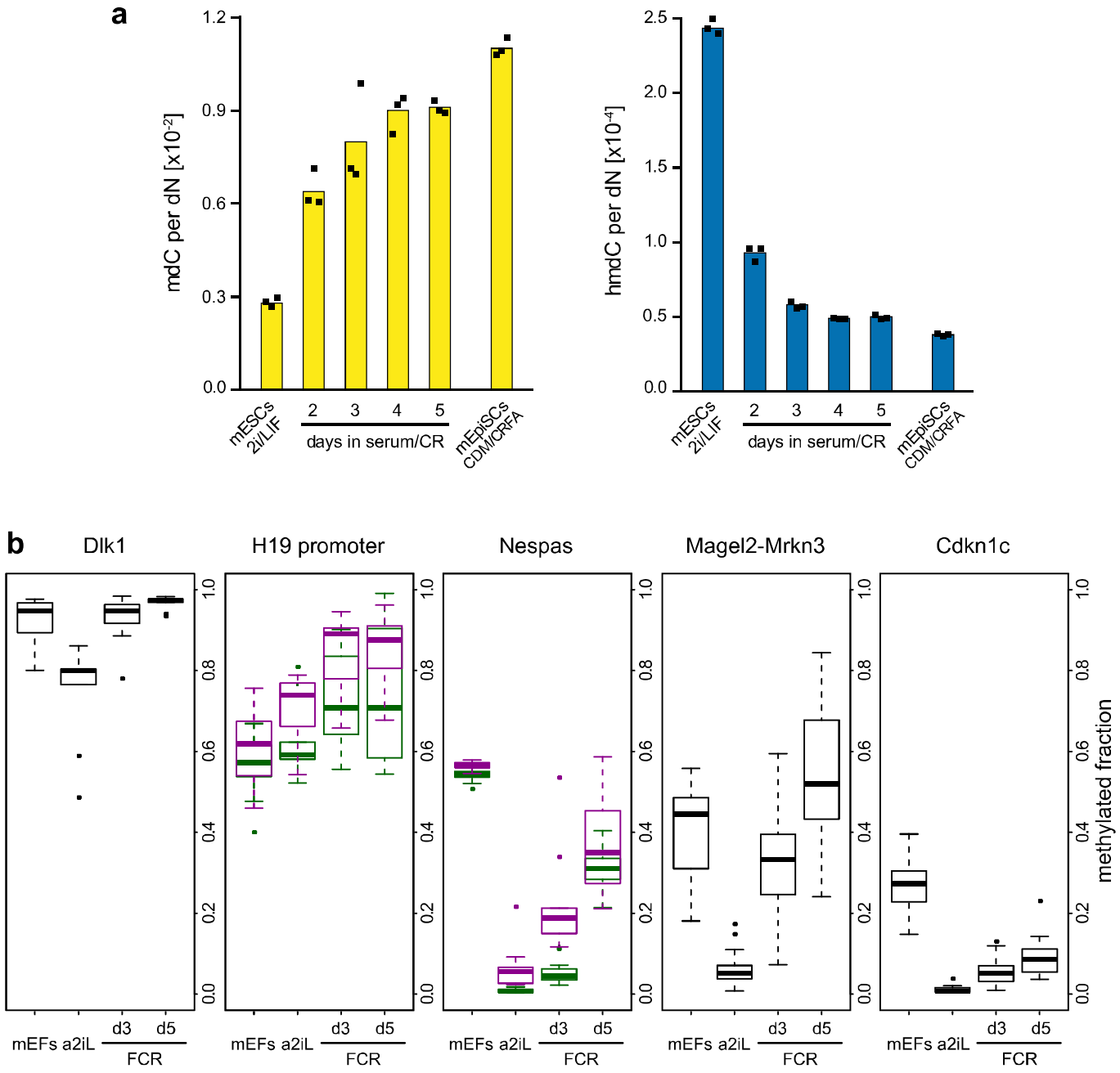
Supplementary information is available for this paper at https://doi.org/10.1038/s41589-020-0621-y.

Correspondence and requests for materials should be addressed to F.S. or T.C.

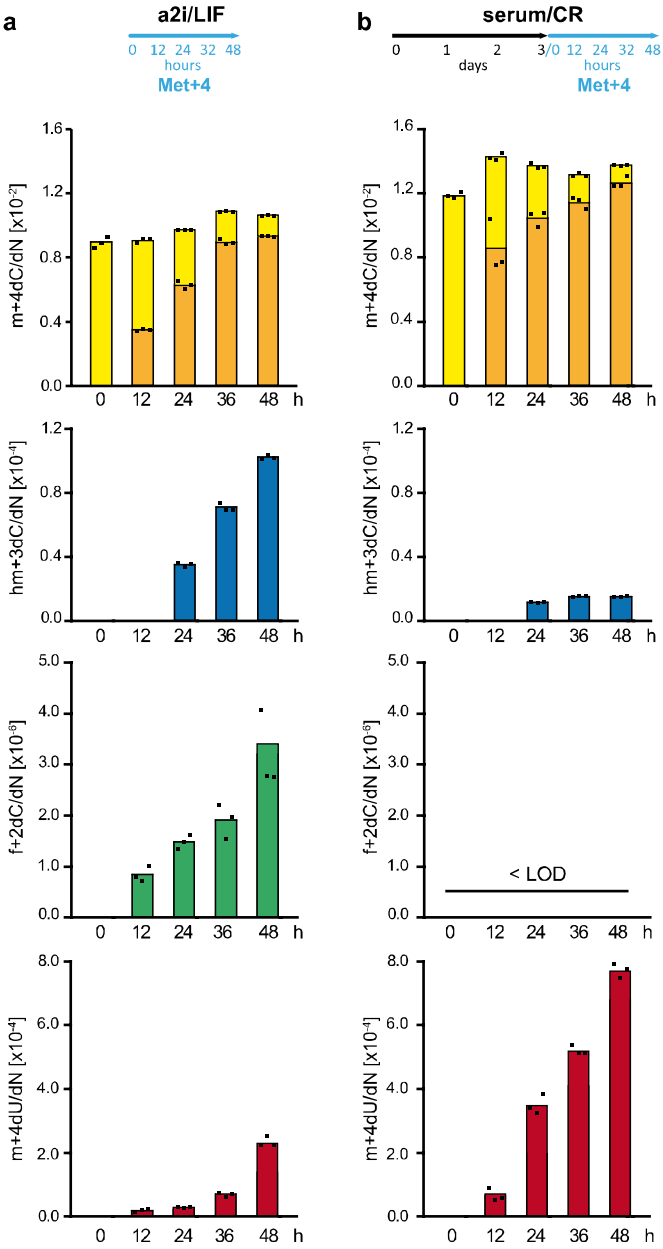
Reprints and permissions information is available at www.nature.com/reprints.



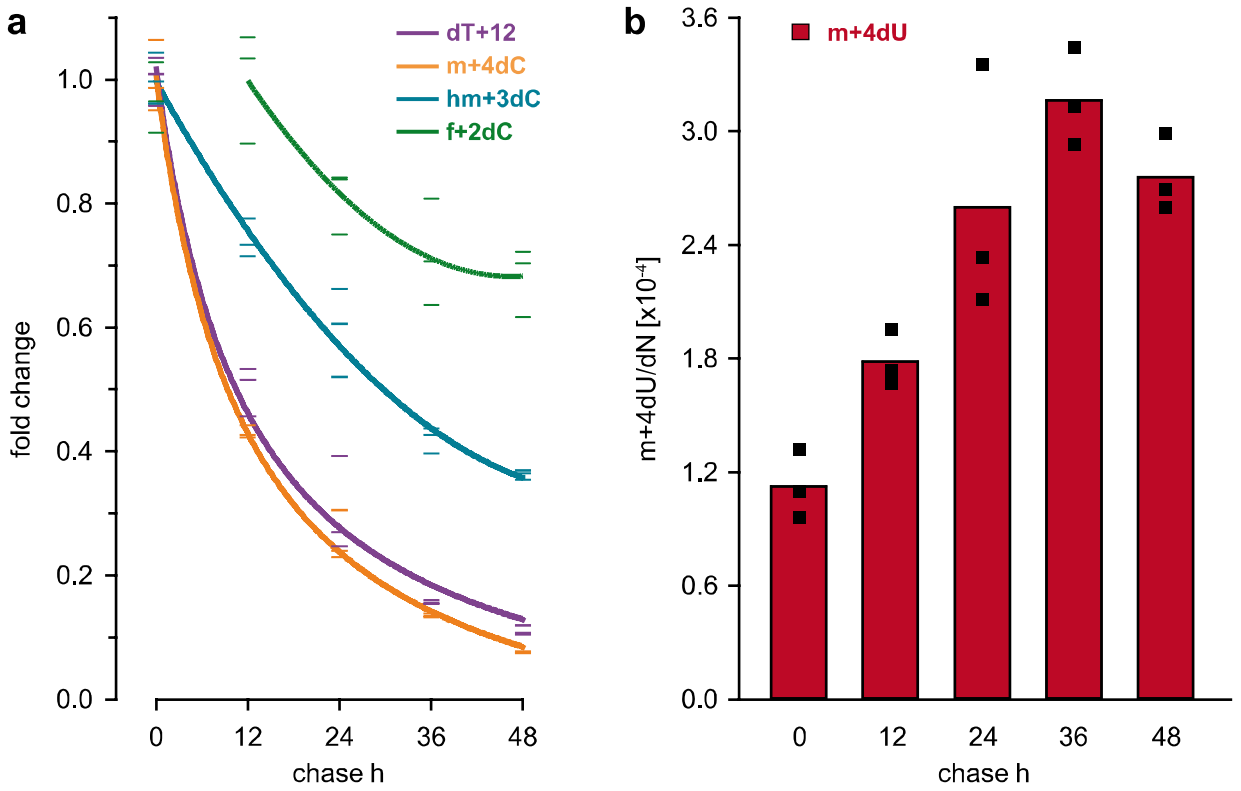
Extended Data Fig. 1 | Expression of pluripotency and early lineage specification factors in mPSCs under different culture conditions. a, YFP fluorescence and bright field images of an Oct4 YFP reporter mPSC line after adaptation to naïve conditions in chemically defined medium (CDM/2i/LIF; left), followed by six days of priming under serum/CHIR/IWR-1 conditions (serum/CR; middle) or long term priming as mEpiSCs in chemically defined medium containing CHIR/IWR-1/FGF 2/activin A (CDM/CRFA; right). Scale bars = 100 µm. **b**, Transcript levels of pluripotency and early lineage specification factors assayed by reverse transcription-qPCR from cultures of the Oct4 YFP mPSC line shown in a under CDM/2i/LIF (naïve), serum/a2i/LIF as well as both serum/CR and CDM/CRFA primed conditions. Note that under serum/a2i/LIF conditions relatively high and low levels of naïve and primed markers are expressed, respectively, while the levels of Oct4 and Nanog transcripts are similar to those in naïve conditions. Also, transcript levels of naïve and primed pluripotency factors are similar under short term serum/CR priming conditions and long term primed EpiSCs in CDM/CRFA. Mean (bars) and individual values of fold change Log (diamonds) from three technical replicates are shown. ND = not detectable. Gapdh was used as housekeeping gene reference to calculate ΔCt values.



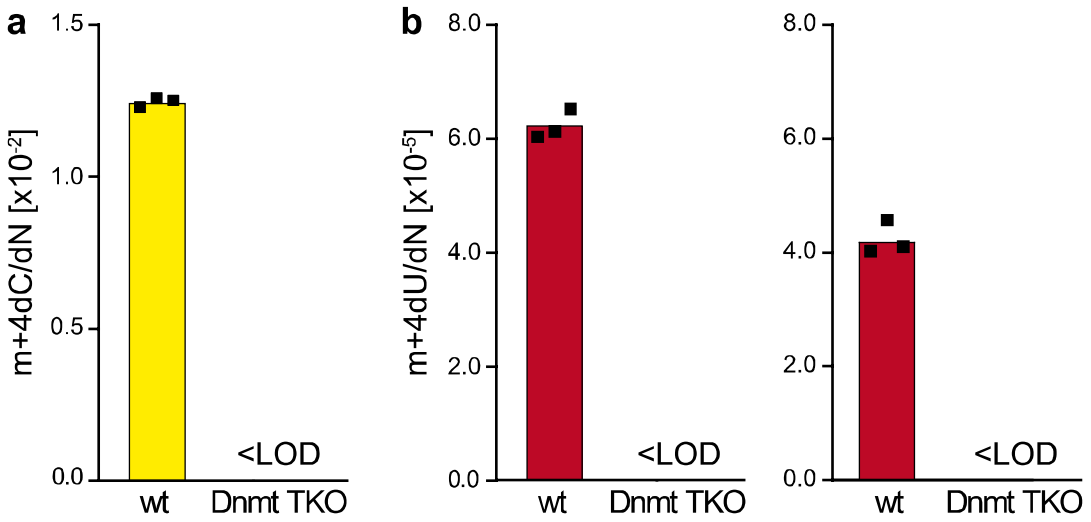
Extended Data Fig. 2 | Global and local transitions of genomic dC modification levels upon priming of mPSCs. a, Global levels of genomic mdC (left panel) and hmdC (right panel) assayed by UHPLC-MS^2 in wt J1 mPSCs transitioning from naïve state (serum/2i/LIF; day 0) to primed state over five days in serum/CHIR/IWR-1 (serum/CR) and in the Oct4-YFP reporter mEpiSC line derived and permanently cultured in serum free CDM containing CHIR/IWR-1/FGF 2/activin A (CRFA). Mean (bars) and individual values (black squares) from three technical replicates are shown. **b**, High depth targeted amplicon bisulfite sequencing analysis of selected secondary imprints in mPSCs upon transition from a2i/LIF (a2iL) to serum/CR (FCR) conditions. Primary mouse embryonic fibroblasts (mEFs) are shown as reference for differentiated primary somatic cells. CpG methylation levels at individual positions were averaged over each amplicon. Size, numbers of CpG sites and genomic coordinates are reported in Supplementary Table 2. Boxes display the inter-quartile range (IQR), where lower and upper edges represent the first and third quartile, respectively, and the thicker horizontal line is the median. Whiskers mark 1.5xIQR and dots represent outliers beyond 1.5xIQR. Magenta and green box plots represent distinct amplicons within the H19 promoter region and Nespas. Results from individual culture samples are shown. Numbers of aligned reads used for quantifications (range depends on sample and CpG position): Dlk1: 5.5k-23k; H19: green 13k-35K, purple 22k-39k; Nespas: green 26k-67k; purple 32k-80k; Magel-Mrkn3: 14k-36.7k; Cdkn1c: 10k-26.7k.



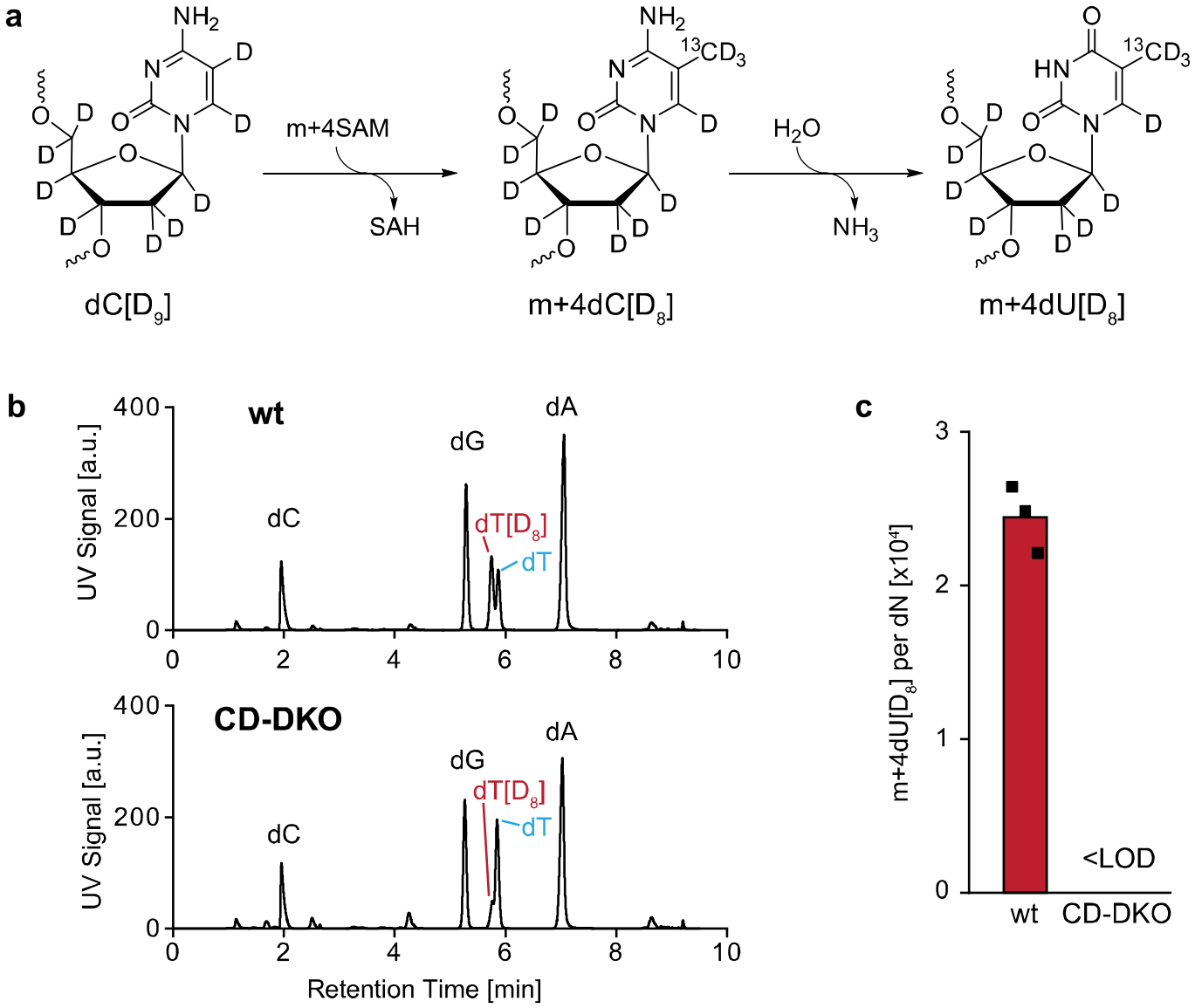
Extended Data Fig. 3 | Drastic reduction of global genomic mdC oxidation and increase of mdC-to-mdU conversion upon transition to primed pluripotency (related to Fig. 1). Time course analysis of genomic dC derivatives upon metabolic labelling with m+4Met under naïve (**a**; a2i/LIF) and priming (**b**; serum/CR) conditions. Independent biological replicate of the experiment shown in Fig. 1c,d. The labelling/time course schedules are shown at the top. Global levels of unlabeled mC (pale yellow) and m+4dC (dark yellow), hm+3dC (blue) and f+2dC (green) are shown as mean (bars) and individual values (black squares) of three technical replicates. LOD = limt of detection.



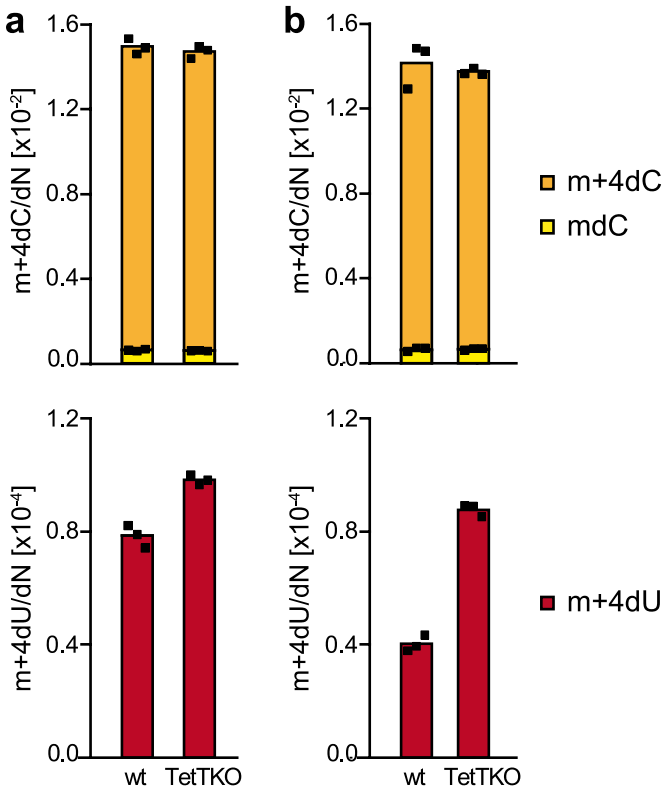
Extended Data Fig. 4 | Turnover of genomic mdC and its derivatives under priming conditions (related to Fig. 2). a, **b**, Independent biological replicate of the experiment reported in Fig. 2c,e. **a**, Global profiles of the indicated labelled cytosine derivatives and dT+12 in the genome of mPSCs under priming conditions (serum/CR) upon chasing of m+4Met. Each panel shows fold change values relative to the highest level for each modification from an individual experiment. Dashes represent single technical replicate measurements (n = 3) from single samples. Full lines lines represent first order decay curves fitted to mean values (in the case of f+2dC t = 0 h was excluded from curve fitting). **b**, Absolute levels of genomic m+4Met from the same samples as in a. Black squares represent single technical replicate measurements (n = 3) from single samples and bars show their mean values.



Extended Data Fig. 5 | Genomic m+4dU is generated exclusively from genomic m+4dC. Global levels of m+4dC (**a**; yellow) and m+4dU (**b**; red) in the genome of wt and Dnmt TKO J1 mPSCs after priming for five days in the presence m+4Met. In Dnmt TKO cells neither m+4dC nor m+4dU were above background levels (LOD = limit of detection). The two panels in b show results from two independent replicate cultures (n = 2). Mean (bars) and individual values (black squares) of three technical replicates are shown.



Extended Data Fig. 6 | Double labelling of primed wt and CD DKO J1 with m+4Met and dC[D9] (related to Fig. 4). a, Conversion of dC[D9] into m+4dC[D8] and potentially m + 4dU[D8] within gDNA. SAH = S adenosylhomocysteine. **b**, UHPLC chromatogram (UV trace) showing the retention delay of dT[D8] relative to natural dT. **c**, Quantification of m+4dU[D8] accumulated in wt and CD DKO PSCs upon priming in the presence of m+4Met and dC[D8]. Mean (bars) and individual values (black squares) of three technical replicates are shown.



Extended Data Fig. 7 | Tet proteins do not trigger DNA repair events that contribute to turnover of mdC into dT (related to Fig. 3). Global levels of mdC, m+4dC and m+4dU (red) in the genome of parental (Tet1-3 proficient) and Tet TKO E14tg2a mPSCs after priming for five days in the presence m+4Met. **a**) and **b**) show independent biological replicates. Mean (bars) and individual values (black squares) of three technical replicates are shown.

3.6 ^{1}H NMR chemical exchange techniques reveal local and global effects of oxidized cytosine derivatives

R. C. A. Dubini, E. Korytiaková, T. Schinkel, P. Heinrichs, T. Carell, P. Rovó, *ACS Phys. Chem Au* **2022**, https://doi.org/10.1021/acsphyschemau.1c00050

Prolog

5-Carboxy-2'-desoxycytidin ist eine relativ selten vorkommende epigenetisch relevante Modifikation die in genomischer DNA auftritt und eine bisher nicht vollständig aufgeklärte Funktion im Genom einnimmt. Im Rahmen dieser Arbeit wurde untersucht, ob das Vorhandensein der Carboxylfunktionalität im genomischen Kontext einen Einfluss auf die Struktur der doppelstrangigen DNA ausübt. Hierzu wurde ein synthetisches cadC-enthaltendes palindromisches Oligonukleotid als Modellsystem über Festphasensynthese hergestellt. Um den Einfluss der Modifikation auf die physikalischen Eigenschaften von doppelsträngiger DNA zu untersuchen, wurde das aus 12 Nukleobasen bestehende Oligonukleotid mittels ^{1}H-NMR Spektroskopie analysiert. Hierbei wurde die Dynamik des Systems auf atomarer Ebene bei unterschiedlichen Temperaturen und pH-Werten untersucht. Die spektroskopischen Daten zeigten keinen Effekt der Modifikation auf die B-DNA Struktur des Doppelstrangs, jedoch konnte eine eindeutige Destabilisierung des Doppelstrangs bei niedrigeren pH-Werten (pH 5.8 und pH 4.7) nachgewiesen werden.

Autorenbeitrag

In dieser Arbeit war ich für die Synthese des cadC Phosphoramidits und dessen Einbau in das zu untersuchende Oligonukleotid zuständig. Nach der Darstellung des Oligonukleotids per Festphasensynthese habe ich die anschließenden Entschützungen und Aufreinigungen durchgeführt. Außerdem wurde das Schmelzverhalten des Oligonukleotids per UV/Vis-Messungen von mir durchgeführt. *Romeo Dubini* führte die NMR-spektroskopischen Messungen des Oligonukleotids bei verschiedenen pH-Werten, sowie die Auswertung der erhaltenen Daten durch.

ACS PHYSICAL CHEMISTRY Au
AN OPEN ACCESS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/physchemau Article

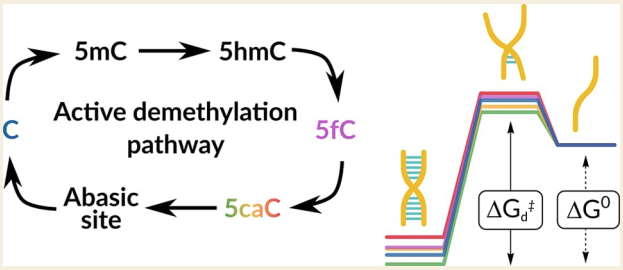
^{1}H NMR Chemical Exchange Techniques Reveal Local and Global Effects of Oxidized Cytosine Derivatives

Romeo C. A. Dubini, Eva Korytiaková, Thea Schinkel,[§] Pia Heinrichs,[§] Thomas Carell, and Petra Rovó*

Cite This: https://doi.org/10.1021/acsphyschemau.1c00050 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: 5-Carboxycytosine (5caC) is a rare epigenetic modification found in nucleic acids of all domains of life. Despite its sparse genomic abundance, 5caC is presumed to play essential regulatory roles in transcription, maintenance and base-excision processes in DNA. In this work, we utilize nuclear magnetic resonance (NMR) spectroscopy to address the effects of 5caC incorporation into canonical DNA strands at multiple pH and temperature conditions. Our results demonstrate that 5caC has a pH-dependent global destabilizing and a base-pair mobility enhancing local impact on dsDNA, albeit without any detectable influence on the ground-state B-DNA structure. Measurement of hybridization thermodynamics and kinetics of 5caC-bearing DNA duplexes highlighted how acidic environment (pH 5.8 and 4.7) destabilizes the double-stranded structure by ∼10−20 kJ mol^{-1} at 37 °C when compared to the same sample at neutral pH. Protonation of 5caC results in a lower activation energy for the dissociation process and a higher barrier for annealing. Studies on conformational exchange on the microsecond time scale regime revealed a sharply localized base-pair motion involving exclusively the modified site and its immediate surroundings. By direct comparison with canonical and 5-formylcytosine (5fC)-edited strands, we were able to address the impact of the two most oxidized naturally occurring cytosine derivatives in the genome. These insights on 5caC's subtle sensitivity to acidic pH contribute to the long-standing questions of its capacity as a substrate in base excision repair processes and its purpose as an independent, stable epigenetic mark.



KEYWORDS: *DNA, Epigenetic modifications, 5-formylcytosine, 5-carboxycytosine, NMR, Dynamics*

Downloaded via 77.2.118.111 on April 18, 2022 at 17:54:06 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

INTRODUCTION

Apart from the four canonical bases, DNA may also contain modified versions of cytosine, thymine, or adenosine nucleosides.[1,2] Such naturally occurring DNA modifications are broadly named epigenetically modified bases, and they constitute an additional regulatory layer by extending genomic complexity and have been shown to play several crucial roles.[3] The discovery of ten-eleven translocation (TET)-induced oxidation of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) has been a key catalyst for the exploration and subsequent characterization of both 5-formylcytosine (5fC) and 5-carboxycytosine (5caC).[4]

5fC and 5caC retain a yet undefined number of functional roles, besides being intermediates within the active demethylation pathway,[5,6] as they represent genomically stable, semipermanent modifications with clearly defined tissue distributional patterns.[7−9] Both have been reported as abnormally abundant in prostate, breast and plasma cells cancer.[10−13] Their biological significance is not limited to the initial appearance and progression of diseases, since 5fC and 5caC are transiently accumulated during lineage specification of neural stem cells (NSCs) in culture and *in vivo*,[14] reduce the rate and substrate specificity of RNA polymerase II transcription,[15] or can be selectively recognized by specialized proteins.[16]

Despite extensive research efforts in recent years, it is yet unclear how reader proteins recognize 5fC and 5caC with high specificity and selectivity regardless of their sparse genomic abundance and their chemical similarity to canonical, methylated or hydroxymethylated cytosines. 5fC has been more extensively studied compared to its carboxylated counterpart. Recent research endeavors clarified initial contradictory reports about its impact on structure and stability, leaning toward the notion that 5fC does not affect the B-DNA form.[17−19] As a result of the destabilizing effect that formylation imparts on the C−G base pair, 5fC was found to facilitate melting and hinders annealing, although without affecting the structure to any measurable extent, an aspect which has been investigated via FT-IR and NMR spectroscopy.[20,21]

Received: December 13, 2021
Revised: January 25, 2022
Accepted: January 26, 2022

ACS Publications
© XXXX The Authors. Published by American Chemical Society

From a structural perspective, it has been recently suggested that 5caC is able to induce a kink in dsDNA, and such geometric alteration has been deemed essential for recognition and enzymatic action by Thymine DNA Glycosylase (TDG).[22] On the other hand, disagreement arises concerning its impact on dsDNA melting and annealing, with studies reporting inconsistent de/stabilization-related properties even when studied under nearly identical conditions.[23−25] While 5fC-induced weakening of the base pair was found to be independent of the mildly basic or acidic conditions naturally occurring within different tissues and/or organisms, with the introduction of the titrable carboxyl group, a pH-dependent behavior emerges for 5caC.[25,26] Protonation/deprotonation events of the carboxyl group induce 5caC to act as an electron-withdrawing group (EWG)/electron-donating group (EDG), respectively. These aspects have been reported to be biologically significant. Indeed, the carboxyl group protonation state has a subtle impact on hydrogen bonding, a behavior that was rationalized via the pK_a values of the two solvent-exposed sites: the nitrogen atom N3 and the carboxyl group itself.[21,26,27] Correspondingly, activity studies on DNA polymerases also concluded that 5caC acts as a base-pair mismatch during DNA replication, signaling that protonated 5caC is a highly destabilizing entity in the context of dsDNA.[28] In addition, structural biology investigations have suggested that the degree of protonation of 5caC's carboxyl group might be a key factor in the mechanism of excision operated by TDG.[26]

Even though other studies have considerably advanced our understanding of epigenetically modified DNA bases within the context of nucleic acids, a number of aspects remain unsettled. In this paper, we seek to shed light on the extent to which 5caC-edited dsDNA's structural features and kinetics-related phenomena deviate from its formylated and canonical equivalents. By employing state-of-the-art methodologies and analytical frameworks in solution-state NMR spectroscopy, we aimed at providing a noninvasive, label-free and site-specific description of the structure and dynamics of these samples using three distinct approaches. First, we compared the impact of the 5caC modification on dsDNA at pH 7.0, 5.8, and 4.7 with respect to 5fC and canonical C by measuring ^{1}H, ^{13}C, and ^{15}N chemical shift perturbations. Second, we applied temperature-dependent ^{1}H chemical exchange saturation transfer (CEST) experiments to assess the impact of 5caC on dsDNA melting and annealing processes providing site-specific parameters for dissociation ($\Delta G^{\ddagger}_{d,37°C}$, $\Delta H^{\ddagger}_{d}$, $\Delta S^{\ddagger}_{d}$) and association kinetics ($\Delta G^{\ddagger}_{a,37°C}$, $\Delta H^{\ddagger}_{a}$, $\Delta S^{\ddagger}_{a}$) and for melting thermodynamics ($\Delta G^{\circ}_{37°C}$, $\Delta H°$, $\Delta S°$).[20,29] Third, we explored the microsecond time scale exchange kinetics via ^{1}H on-resonance $R_{1\rho}$ relaxation dispersion (RD) targeting potential base-specific motions. Watson−Crick (WC) to Hoogsteen (HG) base-pair exchange processes occurring in this regime have been characterized in the context of canonical DNA and protein−DNA complexes,[30−32] but no other exchange phenomena are known to occur in undamaged, non-mismatched DNA helices which are not interacting with a reader protein or enzyme.[33,34]

Collectively, our findings reveal that 5caC's pH-induced chameleonic behavior is not due to any permanent structural changes, opposite to previously reported results.[22] In fact, we observed that the repercussions of 5caC and 5fC incorporation into otherwise canonical dsDNA are uniquely perceptible in the context of conformational dynamics affecting at least two distinct time scales: a slower one (tens of milliseconds) and a faster one (hundreds of microseconds). In addition, we found that carboxycytosine does noticeably affect dsDNA annealing and melting phenomena when exposed to progressively lower pH conditions. Our unified analysis provides a highly comparable and comprehensive overview of the mechanistic details involving both global (such as energetics of DNA strands association and dissociation) and local (site specific motions involving a single base-pair) dynamics phenomena, and sets the stage for future studies of protein−DNA interactions.

RESULTS

We considered the pH-dependent structural features, thermodynamic stability, and site-specific dynamics of a self-complementary 12mer homocarboxylated DNA with the sequence of 5′-GCGATXGATCGC-3′ where X stands for 5caC (Figure 1). The corresponding samples were named

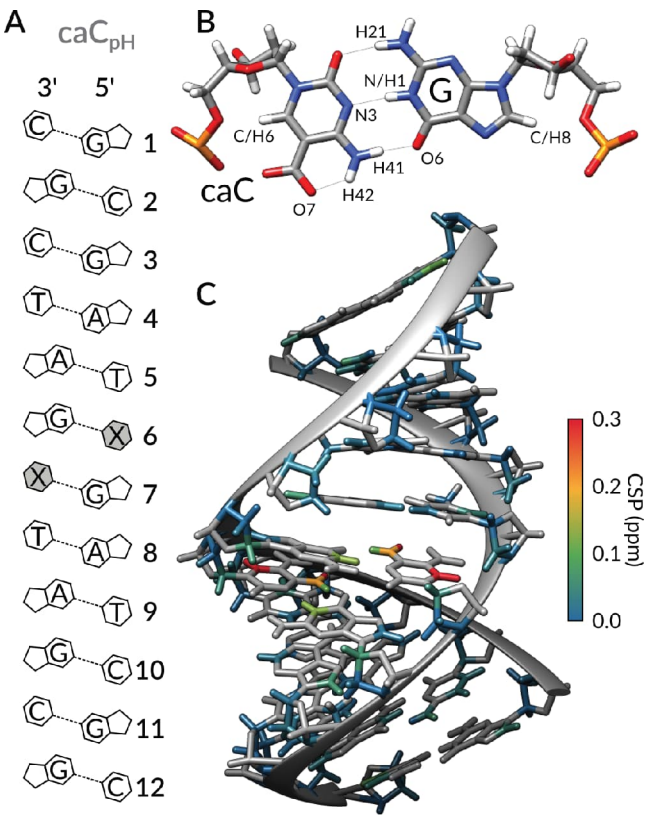


Figure 1. (A) DNA sequence of caC$_{pH}$'s model sequence. The modified cytosine nucleoside is highlighted in gray, marked as X. (B) Structural model showing the expected 5caC−G base pair conformation, consistent with ^{1}H, ^{13}C, and ^{15}N chemical shift values. (C) Absolute CSP values comparing samples at pH 7.0 and 4.7 are schematically displayed onto the B-DNA structural model of caC$_{pH}$.

caC$_{7.0}$, caC$_{5.8}$, and caC$_{4.7}$ indicating the pH at which they were studied. In order to evaluate the influence of cytosine carboxylation in a broader context, we compare the chemical shifts and the NMR-derived thermodynamics, kinetics, and dynamics parameters to analogous values obtained for canonical (C$_{7.0}$) and 5fC-modified (fC$_{7.0}$) samples featuring a sequence identical as what we considered here, which were previously reported in ref 20.

https://doi.org/10.1021/acsphyschemau.1c00050
ACS Phys. Chem Au XXXX, XXX, XXX−XXX

Structural Impact by Chemical Shift Analysis

To spot any unusual features induced by inclusion of carboxylated cytosine within otherwise canonical DNA, we recorded a comprehensive set of homo- (^{1}H−^{1}H NOESY) and heteronuclear (^{1}H−^{13}C HSQC, ^{1}H−^{15}N SOFAST HMQC) 2D spectra for resonance assignment and chemical shift analysis purposes.[35]

Previous crystallographic studies have reported contrasting results concerning a potential geometric alteration of the dsDNA helix induced by 5caC.[22,23] Our results, in support of ref 23, indicate that the 5caC nucleobase does not induce any detectable, permanent deviation from the canonical B-DNA structure. At pH 7.0, when compared to the 5fC−G base-pairing interactions, 5caC−G appears to be equally well-tolerated within a canonical double-helix architecture: all of ^{1}H−^{13}C and ^{1}H−^{15}N cross-peaks are superimposable among the canonical, and 5fC and 5caC-containing 12mer DNA constructs, with the sole exception represented by those nuclei in direct proximity to the epigenetic modification (Figures 2, S1, and S2).

An analogous comparison of $caC_{7.0}$ with $caC_{5.8}$ and $caC_{4.7}$ points to no detectable structural change upon acidification. ^{1}H, ^{13}C, and ^{15}N resonances, $^3J_{HH}$ couplings and ^{1}H−^{1}H NOESY cross-peak patterns suggest that carboxylated cytosine leaves the B-DNA structure entirely unperturbed at both neutral and acidic pH values, resulting in no substantial structural difference from the canonical or formylated cytosine-bearing constructs in any studied condition (Figures S3−S6).

Figure 2 compares the perturbations of representative ^{1}H, ^{13}C, and ^{15}N chemical shifts of $fC_{7.0}$, $caC_{7.0}$, $caC_{5.8}$, and $caC_{4.7}$ with respect to the shifts of the canonical 12mer sequence. Figure 2A shows imino protons, which are sensitive reporters of the strength of base-pairing: downfield (upfield) shifted signals, appearing at higher (lower) ppm values, revealing stronger (weaker) intermolecular hydrogen bonds. Interestingly, the only responsive site to pH changes is G7H1, which reports on the X:G base pair, where X is either C, 5fC, or 5caC at the three distinct pH conditions we selected. Substantial chemical shift perturbations are also visible for aromatic ^{1}H sites T5H6 and X6H6, ^{13}C X6C6, and ^{15}N G7N1 (Figures 2B−D and S7−S10). Among those nuclei, the trends suggest that $caC_{4.7}$'s resonances resemble $fC_{7.0}$ the most, while in neutral conditions $caC_{7.0}$ is comparable with canonical $C_{7.0}$.

These pH-induced shifts were most prominent for H-bond forming protons, namely, for caC6H41, H42, and G7H1. The nature of the effect is related to the reshuffling of electron densities around the base-pairing atoms due to altered Pauli repulsion between occupied atomic orbitals.[36] An increased ^{1}H chemical shift is associated with a decreased electron density (^{1}H shielding) around the proton and hence with a stronger H-bond, or shorter H−X distance. As the carboxyl group of 5caC becomes progressively protonated with decreasing pH, the intrabase H-bond between caC6H42 and the carboxyl oxygen O7 gets weaker (caC6H42 $\Delta\delta$ = −0.3 ppm), while the interbase H-bond formed by caC6H41 and G7O6 (Figure 1B) gets stronger (caC6H41 $\Delta\delta$ = +0.1 ppm). Meanwhile, hydrogen bonding between G7H1 and caC6N3 weakens (G7H1 $\Delta\delta$ = −0.15 ppm) due to the decreased electron density at N3 caused by the EWG properties of the protonated carboxyl group, compensating for the increased base-pairing stability gained by the H41−O6 interaction (Figures S7−S10). The fine balance between the strengths of the intra- and interbases hydrogen bonds in the 5caC−G base pair leads to close to optimal base-pairing both at neutral and mildly acidic conditions, leaving the B-DNA structure unperturbed through the entire studied pH range.

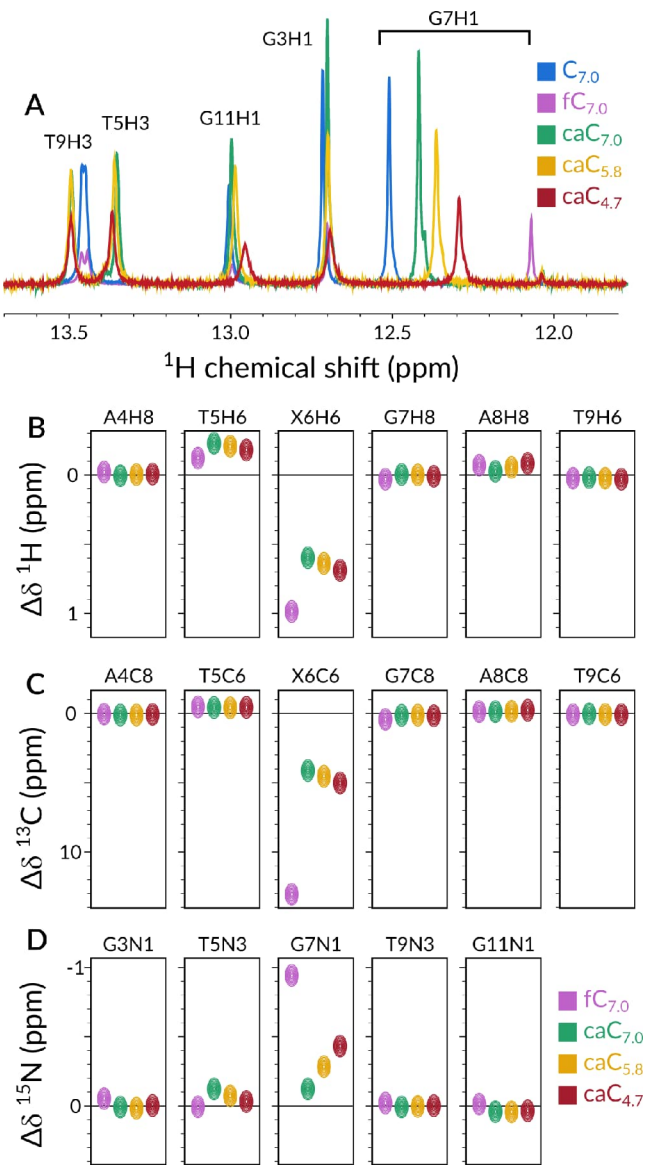


Figure 2. Comparison of chemical shift perturbations $\Delta\delta$ for representative imino ^{1}H (A), aromatic ^{1}H (B), aromatic ^{13}C (C), and imino ^{15}N (D) resonances with respect to the chemical shifts of the canonical 12mer sample (blue spectrum in A, black horizontal line in B−D). Magenta, green, yellow, and red symbols represent $fC_{7.0}$, $caC_{7.0}$, $caC_{5.8}$, and $caC_{4.7}$, respectively. A full comparison of all comparable chemical shifts between the five samples is displayed in Figures S7−S10.

5caC's Influence on DNA Melting and Annealing

A thorough characterization of nucleic acids' global structural rearrangements, such as folding, melting, annealing, and binding should entail a comprehensive analysis of kinetic events occurring on the millisecond to second time scale.[37] CEST experiments have found adoption in modern biomolecular NMR, allowing for quantitative and site-specific determination of population, chemical shift, and exchange kinetics of sparsely populated conformations.[29,38] When measured in a temperature-dependent fashion, the shift in

https://doi.org/10.1021/acsphyschemau.1c00050
ACS Phys. Chem Au XXXX, XXX, XXX−XXX

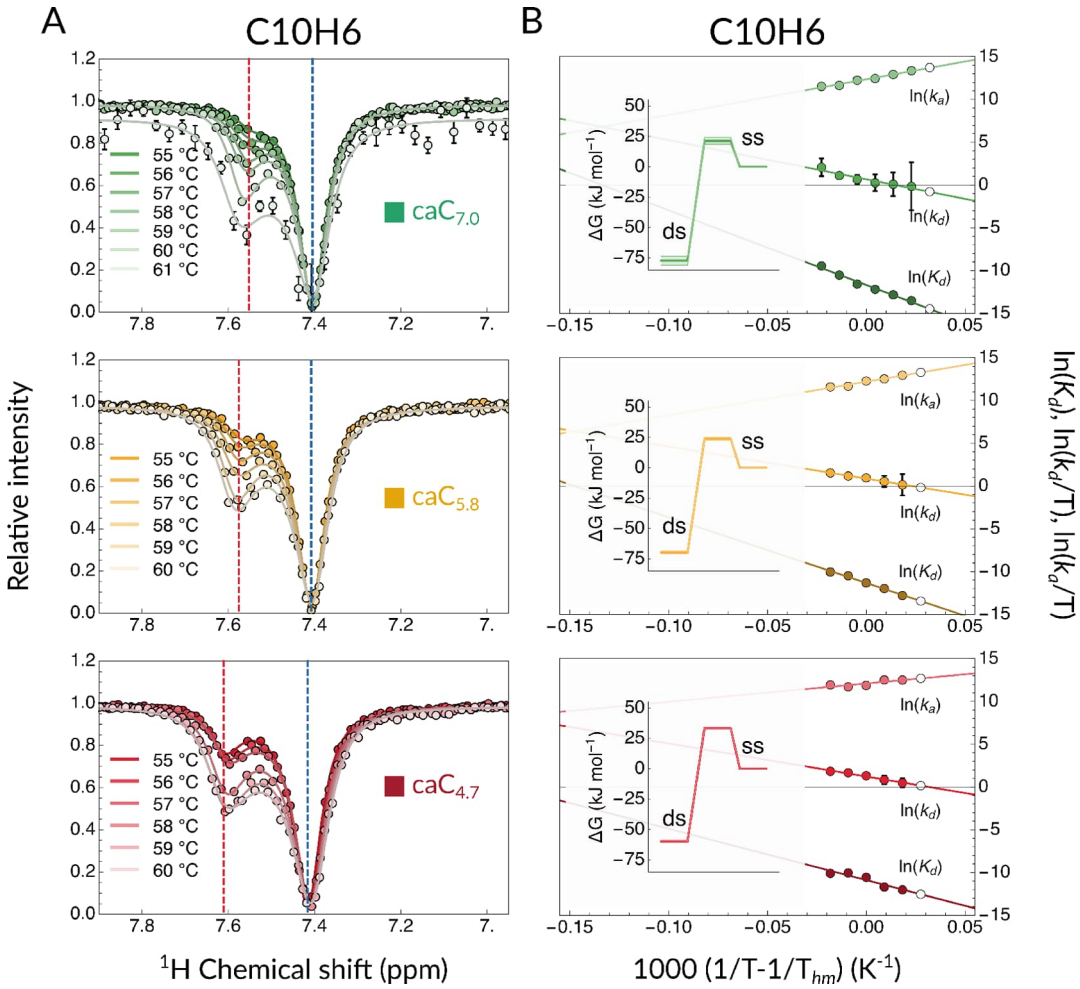


Figure 3. (A) Temperature-dependent CEST melting profiles together with the obtained two-site exchange fits for proton C10H6. $caC_{7.0}$, $caC_{5.8}$, and $caC_{4.7}$ are shown in shades of green, yellow, and red, respectively. Dashed blue and red lines indicate dsDNA and ssDNA chemical shift values. (B) van't Hoff plots relative to the CEST melting profiles. Shades of green, yellow, and red indicate data entries and linear fits of ln k_a, ln k_d, and ln K_d vs $1/T - 1/T_{hm}$, where T_{hm} is the harmonic mean of the measured temperatures, for $caC_{7.0}$, $caC_{5.8}$, and $caC_{4.7}$, respectively. White data points represent back-calculated values for 55 °C that were used to validate the CEST fits. Insets present the relevant Gibbs free energy plots at each pH condition at 37 °C.

exchange parameters can reveal atomistic details about the melting thermodynamics and kinetics of the studied system providing unprecedented insights into the molecular processes. In pursuance of the study of 5caC-induced DNA destabilization, we proceeded by recording CEST profiles for the aromatic protons in all caC_{pH} samples in the 55−61 °C range. As an example, Figure 3A displays the melting CEST profiles for C10H6 at three pH values (profiles for other comparable protons can be found in Figures S11−S18). The appearance of a distinctive secondary dip at increasingly higher temperatures indicates the presence of an alternative conformer, which we identify as the single-stranded conformation (ssDNA) as per comparative chemical shift analysis.

To obtain a more quantitative comparison, we fit the CEST profiles to a two-state exchange model (dsDNA ⇌ 2 ssDNA) that yielded a numerical estimation of populations (p_D for dsDNA, $1 - p_D$ for ssDNA), exchange kinetics (k_{ex}), and chemical shifts of the exchanging states at each of the highest temperatures, while we used the lowest temperature of the ensemble for each sample (55 °C) to compare the predicted back-calculated value from the fits to the experimental data points. ^{1}H longitudinal (R_1) and transverse (R_2) relaxation rates were measured separately at multiple temperatures and used as inputs for the CEST fits assuming that the rate constants of the dsDNA and ssDNA states are the same.

As an example, Figure 3B shows the logarithm of the obtained kinetic rates and equilibrium constants for C10H6. Consistently with results obtained for $fC_{7.0}$ and $C_{7.0}$, two observations can be made: (i) a linear fit could be identified for kinetic rates and equilibrium constants against T^{-1}, suggesting a single transition state is plausible for the melting and annealing phenomena, and (ii) all sites show thermally activated kinetics with a positive dissociation barrier (Arrhenius behavior) and a negative association barrier (anti-Arrhenius).

In addition, CEST measurements run at multiple distinct temperatures allow for the extraction of Gibbs free energy and related parameters by leveraging the temperature-dependent nature of kinetics and thermodynamic phenomena. In order to achieve an estimation of the degree of protonation-induced destabilization to dsDNA, we applied our recently described methodological framework for the decomposition of traditional CEST output into enthalpic and entropic stability and activation changes.[20] Succinctly, we assumed the observed

https://doi.org/10.1021/acsphyschemau.1c00050
ACS Phys. Chem Au XXXX, XXX, XXX−XXX

dynamic equilibrium is the reversible melting/annealing process of a dsDNA strand into two single-stranded DNA sequences. If the concentration of DNA is known, then kinetics of association (k_a) and dissociation (k_d) can be extracted. From the temperature dependence of k_a and k_d, activation barriers for both process ($\Delta G^{\ddagger}_{d,37°C}$ and $\Delta G^{\ddagger}_{a,37°C}$) can be derived. Analogously, the ratio between $k_a(T)$ and $k_d(T)$ rates allows for the determination of the equilibrium dissociation constant (K_d) and consequently the quantification of thermodynamic parameters such as $\Delta G^{\circ}_{37°C}$.

This analysis allows us to directly compare $caC_{7.0}$, $caC_{5.8}$, and $caC_{4.7}$ not only between them, but also with otherwise identical canonical and formylated samples ($C_{7.0}$ and $fC_{7.0}$, respectively), which we previously discussed in ref 20. In Table 1, we compare three proton reporters across all five samples

Table 1. Thermodynamic and Kinetic Parameters of the dsDNA Melting Process Obtained from the van't Hoff and Eyring Analysis of the CEST-Derived Exchange Parameters[a]

sample		$\Delta G^{\circ}_{37°C}$ (kJ mol^{-1})	$\Delta G^{\ddagger}_{d,37°C}$ (kJ mol^{-1})	$\Delta G^{\ddagger}_{a,37°C}$ (kJ mol−1)
$C_{7.0}$	C2H6	71.1 ± 1.7	99.2 ± 2.2	28.1 ± 2.1
	T9H6	70.2 ± 1.8	102.2 ± 1.5	31.9 ± 1.5
	C10H6	71.3 ± 1.4	100.3 ± 1.3	29.0 ± 1.3
$caC_{7.0}$	C2H6	73.3 ± 7.4	93.0 ± 4.3	19.7 ± 4.1
	T9H6	75.7 ± 0.8	99.2 ± 0.9	23.4 ± 0.9
	C10H6	77.3 ± 3.6	98.6 ± 3.2	21.3 ± 2.9
$caC_{5.8}$	C2H6	71.1 ± 1.2	94.8 ± 1.4	23.7 ± 1.3
	T9H6	68.0 ± 0.9	96.6 ± 0.9	28.6 ± 1.0
	C10H6	67.4 ± 1.4	92.4 ± 1.2	24.9 ± 1.1
$caC_{4.7}$	C2H6	58.6 ± 0.7	93.1 ± 0.7	34.5 ± 0.9
	T9H6	62.1 ± 0.4	95.2 ± 0.4	33.1 ± 0.5
	C10H6	56.8 ± 0.6	92.9 ± 0.6	36.1 ± 0.6
$fC_{7.0}$	C2H6	74.7 ± 10.4	99.7 ± 5.6	25. ± 5.5
	T9H6	64.6 ± 1.2	97.1 ± 1.0	32.5 ± 1.0
	C10H6	66.4 ± 1.2	97.5 ± 1.0	31.1 ± 1.1

[a]Errors are given as one standard deviation. An extended version of Table 1 can be found in the Supporting Information file (Table S7).

(comprehensive tables including fitting results for all proton reporters can be found in Tables S1−S6, while an extended version of Table 1 is available in Table S7). From a thermodynamic perspective, among the C-modified samples, $caC_{7.0}$ scores as the most stable one, which is highly comparable to $C_{7.0}$, as observed by UV/vis spectroscopy (Figure S19) as well as in other studies.[23,27] Acidification of the buffer to pH 5.8 destabilizes the double-stranded conformer by ∼2−10 kJ mol^{-1} and further by another ∼6−13 kJ mol^{-1} when the pH is decreased from 5.8 to 4.7. Interestingly, $\Delta G^{\circ}_{37°C}$ for $caC_{5.8}$'s proton reporters are very similar to those we obtained for $fC_{7.0}$. Kinetics of dissociation data, as expected, support the notion that $caC_{7.0}$ and $C_{7.0}$ require the most energy for undergoing a dsDNA → 2 ssDNA conformational transition. According to this metric, $fC_{7.0}$ is slower in undergoing the melting process when compared to $caC_{5.8}$ and $caC_{4.7}$, as $\Delta G^{\ddagger}_{d,37°C}$ is higher by ∼3−7 kJ mol^{-1}. Lastly, kinetics of association suggest that $caC_{4.7}$ is the slowest in performing an annealing process, followed by $fC_{7.0}$. Compellingly, $\Delta G^{\ddagger}_{a,37°C}$ data indicate that the rate of association of $caC_{7.0}$ is ∼8 kJ mol^{-1} less energetically demanding when compared to $C_{7.0}$, a behavior that can be rationalized considering the EDG nature of the carboxylate substituent (5caC at neutral pH) when compared to a proton (canonical C).

In Figure 4, we show the correlation between the dissociation and equilibrium free energy changes across all

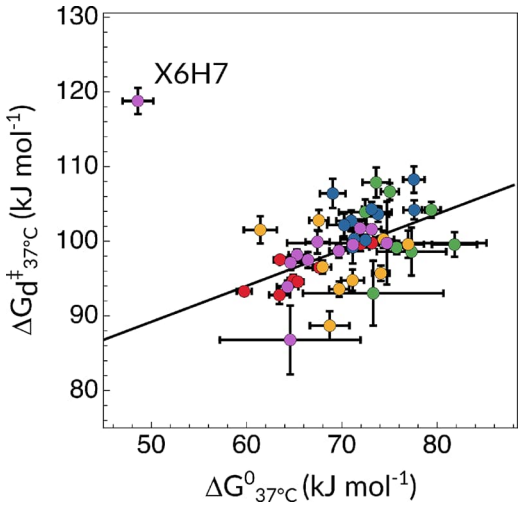


Figure 4. Correlation plot between the changes of the Gibbs free energy of activation for the dissociation process and the equilibrium free energy of previously reported values for $fC_{7.0}$ and $C_{7.0}$ (magenta and blue, respectively) together with 5caC-containing samples ($caC_{7.0}$ as green, $caC_{5.8}$ as yellow, and $caC_{4.7}$ as red) at 37 °C. $fC_{7.0}$ X6H7 (featuring much higher activation free energies and lower equilibrium energies than the rest of the molecule) is an outlier due to its stable intramolecular hydrogen-bond between the formyl O7 atom and the adjacent amino H42 proton.

samples and conditions, for every proton reporter. Here, $\Delta G^{\ddagger}_{d}$ is plotted as a function of ΔG° for all five samples across all eligible proton reporters. Data points accounting for $caC_{7.0}$ and $C_{7.0}$ tend to cluster at the upper right-hand corner of the plot. Conversely, $\Delta G^{\circ 0}_{37°C}$ and $\Delta G^{\ddagger}_{d,37°C}$ values are substantially decreased whenever $fC_{7.0}$, $caC_{5.8}$, or $caC_{4.7}$ is considered, as elaborated above.

Protonation-Induced Microsecond Dynamics

^{1}H and ^{15}N chemical shift values for G7H1/N1 nuclei, both reporters of the centrally positioned 5caC6-G7 base pair's stability, have evidenced that the protonation state of the exocyclic carboxylic group 5caC has a selective impact on this imino proton resonance (Figure 2A, D), which have long been regarded as key indicators of hydrogen bond strength.[23,39] Because of this observation, we aimed at investigating whether this protonation-induced, localized weakening is accompanied by increased probability of local fast time-scale motions.

Our CEST-based kinetic and thermodynamic analysis has established that $caC_{5.8}$, $caC_{4.7}$, and $fC_{7.0}$ destabilize the double-stranded DNA structure without any apparent static, persistent impact on its helical architecture. In order to investigate the presence of a potentially localized conformational exchange which might contribute to the destabilization, we interrogated the faster, microsecond time scale by applying ^{1}H $R_{1\rho}$ relaxation dispersion (RD) methods.[37,40]

In Figure 5, we show X6H6 (where X = C, 5fC or 5caC, depending on the sample under current consideration) ^{1}H on-resonance $R_{1\rho}$ RD profiles measured at 55 °C, a condition that ensures that the melting process is still rather sparse and

https://doi.org/10.1021/acsphyschemau.1c00050
ACS Phys. Chem Au XXXX, XXX, XXX−XXX

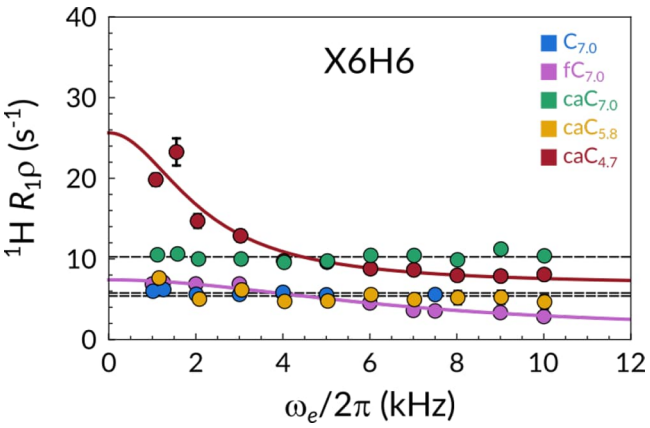


Figure 5. $R_{1\rho}$ relaxation dispersion profiles of $C_{7.0}$ (blue), $fC_{7.0}$ (magenta), $caC_{7.0}$ (green), $caC_{5.8}$ (yellow), and $caC_{4.7}$ (red) recorded at 55 °C. Lines represent best fits to models either accounting for (color-coded) or discounting chemical exchange (black).

infrequent. The data sets recorded for $caC_{7.0}$, $caC_{5.8}$, and $C_{7.0}$ are best fit to a no-exchange model, resulting in flat profiles (black lines). Conversely, profiles for $caC_{4.7}$ and $fC_{7.0}$ fit best to the exchange model, exposing a chemical exchange contribution to its R_2 relaxation rate. Such motions, consistent with a τ_{ex} in the order of hundreds of microseconds, fall within the intermediate exchange regime and are approximately 2 orders of magnitude faster compared to the overall melting process as characterized by our CEST measurements. This result is especially interesting when comparing the H6 proton of the modified base in samples $caC_{4.7}$ and $fC_{7.0}$ to any other available ^1H nucleus. No other profile is consistent with an exchange phenomenon on this interval (Figures S20−S24), suggesting that the detected phenomenon is unrelated to the previously reported Watson−Crick to Hoogsteen base-pair exchange process.[30−32] In other words, this motions appears to be sharply localized, leaving all other bases unaffected.

This result can be rationalized by considering our previous speculations on the mechanism by which 5fC (and by extension protonated 5caC) weakens the dsDNA conformer.[20] Deprotonated (or not protonated to a sufficient extent) 5caC proved itself to be either a neutral or even a stabilizing factor in terms of base-pair strength and agrees well both with NOESY-derived chemical shift values and the literature. Instead, whenever the pH of the buffer is sufficiently acidified, the weakening of the 5caC-G hydrogen bond induces a chemical exchange process in the intermediate microsecond time scale. The fact that this motion is sharply localized at the modified site seems to suggest that the protonation of the exocyclic carboxyl moiety and the consequent lowered basicity of the hydrogen-bonded N3 atom could be deemed responsible for generating a sparse and transitory fraying event in the middle of the DNA strand, presumably synergistic with the global destabilization effect on the whole structure.

DISCUSSION

In pursuance of understanding the role of 5caC, we addressed the structural and dynamic features of oligomeric DNA double strands carrying a single version of such modification on each strand. Incorporation of carboxycytosine into DNA is a naturally occurring phenomenon that has been mainly discussed within two biological frameworks: (i) its recently devised (and to date mostly obscure) semipermanent epigenetic role and (ii) as a DNA lesion that undergoes the base excision repair (BER) process.[2−4,41]

Considerations on 5caC Protonation Sites

In their report on divergent mechanisms of enzymatic excision for 5caC and 5fC, Maiti et al. have elaborated how, in a TDG-DNA complex, nitrogen N3 of 5caC (Figure 1B) is likely more basic than the carboxyl group and thus undergoes protonation before the carboxylic exocyclic moiety does.[26] In contrast, despite assigning a pK_{a,COO^-} of 4.7 and $pK_{a,N3}$ of 2.1 for the isolated nucleoside, a distinct infrared spectroscopy and quantum mechanical analysis suggests that, when in the context of a dsDNA strand, the first protonation site is the carboxyl group.[25] Our results support the latter idea. In Figure 2A, we show that no other base pair is affected by the pH change over the entire 7.0−4.7 interval. Indeed, T9H3, T5H3, and G3H1 display negligible chemical shift differences, while G11H1 shifts upfield by ∼0.05 ppm likely due to its proximity to the fraying ends of the oligomeric model system. In sharp contrast, the signal reporting on G7H1 consistently shifts upfield with decreasing pH. While this preference for protonation of the weaker base (i.e., 5caC's carboxylate group over N3) is apparently counterintuitive, we reason that in a dsDNA setting the N3 site is well protected from the solvent environment, both for steric and electrostatic reasons.

5fC and 5caC as Semipermanent Modifications

For a long time, 5fC and 5caC have been mainly considered as transient intermediates within the active demethylation pathway.[42] However, research toward their capacity as standalone epigenetic marks has gained increasingly more traction. For instance, both 5fC and 5caC overlap with H3K4me1 marked regions, associated with active transcription.[43] Also, several developmental and metabolic related genes show 5fC enrichment on promoters before gene upregulation,[44] while 5caC has been reported to transiently accumulate at promoter regions preceding gene expression during lineage specification and differentiation.[45] Lastly, a number of cancerous diseases are correlated with a significant enrichment of such oxidized cytosine epigenetic modifications.[10−13]

From a structural and conformational perspective, both 5fC and 5caC have been reported to induce structural changes localized in the proximity of the modified nucleoside,[22,26] while the pH dependence of 5caC characteristics on dsDNA stability has been previously discussed in the context of short oligomers carrying several clustered modifications.[21,27] In the context of longer (90 bp) DNA strands, Ngo et al. reported a 3-fold enhancement of cyclization rate for 5fC containing strands, while the sample carrying 5caC at pH 8.0 showed no detectable difference when compared to canonical cytosine.[46]

In contrast to refs 22 and 26, our results indicate that, identically to 5fC, 5caC does not induce any permanent structural modification detectable by NMR spectroscopy at any pH condition under consideration (Figures 2 and S2−10). In fact, in analogy to 5fC, 5caC was found to affect dsDNA melting and annealing equilibrium and kinetics, rather than B-DNA average structure. On average, and across all sites, $caC_{5.8}$ and $caC_{4.7}$ resemble $fC_{7.0}$, both thermodynamically and kinetically, when it comes to annealing and melting. In contrast, and consistently with previous cyclization essays and FRET studies, the behavior of deprotonated 5caC is most similar to canonical cytosine.

https://doi.org/10.1021/acsphyschemau.1c00050
ACS Phys. Chem Au XXXX, XXX, XXX−XXX

5mC oxidation derivatives have been shown to accumulate and persist in a relatively stable state in certain biological contexts. For this reason, they have been suggested to carry out additional roles apart from being intermediates in biochemical pathways. In light of the foregoing, we contend that our results could correlate with the relative abundance of 5caC and 5fC in cancerous tissues. How protonation of the exocyclic carboxyl moiety affects the thermodynamics and kinetics of melting and annealing *in vitro* has been discussed in this and previous reports.[20,21,25,27] Our results corroborate the protonation-induced destabilization of the double strand, and we speculate that a similar effect might take place in cancerous tissues, triggered by the low pH environment. Cell proliferation is notoriously accelerated in cancer, and the acidic microenvironment where such diseases thrive is widely recognized as a phenotypic trait, making the presence of the two most oxidized cytosine epigenetic derivatives more than circumstantial.[47,48]

5fC- and 5caC-Driven Enzymatic Recognition

In mammals, active DNA demethylation takes place via an enzymatic tandem of TET and TDG, which together govern the initial stages of the BER pathway. In order to successfully complete the removal of 5mC, the presence of either 5fC or 5caC is ultimately necessary, as they are substrates for TDG, which generates the abasic site.[26] Apart from TET and TDG, several additional proteins are able to selectively recognize, bind, and exert their respective enzymatic activity upon 5fC and 5caC.[12,13,49] The mechanism by which different reader enzymes selectively recognize cytosine's epigenetic modifications has long been established as a crucial theme in chemical biology.[6] Across the proposed enzymatic mechanisms, many rely on a specific residue to initiate the base-extrusion process into the active site. For instance, the "pinch−push−pull" mechanism proposed for TDG leans on Arg275 to promote the breakage of an X:G base pair, where X = T, 5fC, or protonated 5caC.[50,51] Alternative studies suggest that partially extruded nucleotide conformations, which are sparse but naturally occurring events, might play a role in recognition and base excision.[52] This second mechanism of action seems to agree with DNA replication studies, which highlighted how 5caC:G base-pair can behave as a DNA lesion.[28]

Although the data hereby presented is not conclusive, we believe our on-resonance $R_{1\rho}$ RD data assist in providing one more piece of evidence in this complex puzzle. By observing site selective and spontaneous (i.e., not triggered by the enzyme) base-flipping of protonated 5caC:G, especially in the context of a weakened dsDNA strand as evidenced in our CEST analysis, we present further experimental evidence that 5fC and protonated 5caC could indeed act as DNA lesions.[33] We hypothesize that, albeit undetected in our study, such kinetic processes could be present at physiological temperatures and decisively impact enzyme recognition and mode of action.

CONCLUSIONS

In this work, we have considered the impact of 5caC incorporation into a model dsDNA oligomer (caC_{pH}) in three pH conditions, namely, at pH 7.0, 5.8, and 4.7. We obtained chemical shift, melting/annealing, and microsecond conformational exchange data which we could reliably compare with our recent study focusing on 5fC and canonical cytosine. Assignment and chemical shift studies on comparable ^{1}H, ^{13}C, and ^{15}N nuclei have shown that there is no evidence of a permanent structural change: all caC_{pH} samples, together with $C_{7.0}$ and $fC_{7.0}$, are compatible with a standard B-DNA helical arrangement. CEST-derived kinetic and thermodynamic data suggested that the reduced cohesion of the X6:G7 base pair, evidenced by chemical shift studies, affects the extent to which nearby bases are able to cooperatively stabilize one another. $caC_{4.7}$ and $fC_{7.0}$ emerged as the most destabilized samples of the cohort, while $caC_{7.0}$ and $C_{7.0}$ showed remarkably similar properties overall.

$caC_{4.7}$ and $fC_{7.0}$ also revealed a detectable chemical exchange process at or in the proximity of the modified nucleoside X6 on the microsecond time scale. The data hereby presented indicate that 5caC's impact on B-DNA is only evident through the lenses of conformational dynamics, as protonation of the exocyclic carboxyl moiety affects the melting-annealing equilibrium as well as induces sparse and localized microsecond time scale base-pair dynamics. We believe our findings are relevant in the context of several open questions concerning this sparse epigenetic mark. Future investigations may consider expanding our initial exploration of the microsecond time scale by recording off-resonance $R_{1\rho}$ RD experiments or studying protein−DNA interactions featuring isotopic labeled 5fC or 5caC nucleosides to unravel the exact mechanistic details of the interaction between TET, TDG (and other enzymes), and cytosine's oxidized derivatives.

MATERIALS AND METHODS

Sample Preparation

The cadC-phosphoramidite (cadC-PA) and subsequently the modified dsDNA samples caC_{pH} were prepared via phosphoramidite chemistry as previously reported.[53] Solid phase synthesis of oligonucleotides containing cadC was performed on an ABI 394 DNA/RNA synthesizer (Applied Biosystems) using standard DNA synthesis conditions with a cartridge scale of 1 μmol. The phosphoramidites Bz-dA, Ac-dC, iBu-dG, and dT as well as the PS carriers were purchased from LinkTechnologies. For the reaction of the cadC-PA a coupling time of 180 s was applied. The terminal DMT protecting group was cleaved after DNA synthesis on the synthesizer. Basic and acidic deprotection of all oligonucleotides was performed according to literature.[53] Purification of the oligonucleotides was achieved with a HPLC system (Agilent 1260 Infinity II 400 bar pump and a Agilent 1260 Infinity II VWD detecting at 260 nm) applying a buffer system of 0.1 M triethylammonium acetate in water (buffer A) and 0.1 M triethylammonium acetate in 80% aqueous MeCN (buffer B), a gradient of 0%−30% buffer B in 45 min and a flow rate of 5.0 mL/min. As stationary phase Nucleodur columns (250/10 mm, C18ec, 5 μm) from Macherey-Nagel were used. Purified oligonucleotides were analyzed by MALDI-TOF (Bruker Autoflex II). Quantification of oligonucleotides was performed via UV/vis spectroscopy with a NanoDrop ND-1000 spectrophotometer at 260 nm. Samples $caC_{7.0}$ and $caC_{5.8}$ were dissolved in aqueous buffers consisting of 15 mM Na_2HPO_4/NaH_2PO_4 (pH 7.0 and 5.8, respectively), 25 mM NaCl in H_2O. Sample $caC_{4.7}$ was prepared by titrating a 1 M HCl solution into the same buffer described above. The thermal stability of the buffer between room temperature and 60 °C was ascertained by pH-meter measurements. Annealing was performed by heating the dsDNA-containing buffer solution to 90 °C for 5 min and slowly cooling it to 5 °C in approximately 90 min, after which it was allowed to return to room temperature. Then, the NMR sample was prepared with the addition of 0.02% NaN_3, 25 μM DSS and 5% D_2O, resulting in final sample concentrations of ∼0.66 mM for all samples, as determined via UV spectrophotometric measurements at 260 nm using the extinction coefficient calculated via the nearest neighbor approximation.

https://doi.org/10.1021/acsphyschemau.1c00050
ACS Phys. Chem Au XXXX, XXX, XXX−XXX

UV/Vis Spectroscopy

UV/vis melting profiles of the oligonucleotides were measured at 260 nm with a JASCO V-650 UV/vis spectrophotometer between 20 and 85 °C (scanning rate of 1 °C/min), and each sample was measured four times. Samples were placed into 100 μL cuvettes and diluted with the same Na_2HPO_4/NaH_2PO_4, NaCl aqueous buffer as used in the NMR experiment. Before each measurement, a layer of mineral oil was placed on the surface of the sample in order to prevent water evaporation. $caC_{7.0}$ was measured at four concentrations (1.25, 2.50, 5.00, and 10.00 μM), while $fC_{7.0}$ and $C_{7.0}$ were measured as described in ref 20. All concentration values yielded absorption values within the linear range of the spectrometer.

NMR Spectroscopy

All experiments were performed on Bruker Avance III spectrometer operating at a ^{1}H Larmor frequency of 800 MHz (corresponding to a magnetic field of 18.8 T) equipped with a 5 mm triple-resonance cryogenically cooled TCI probe. Standard 2D NOESY (mixing time 250 ms) spectra were recorded at 37 °C for resonance assignment. Natural abundance ^{1}H−^{13}C and ^{1}H−^{15}N HSQC and HMQC spectra were recorded using standard fast-pulsing pulse sequences.[54] Site-selective spin relaxation measurements were performed following the SELOPE scheme; these included ^{1}H CEST, on-resonance ^{1}H $R_{1\rho}$, recorded either with a single spin-lock strength of 10 kHz or as an entire RD profile ranging from 1 to 10 kHz, and ^{1}H R_1 experiments at temperatures between 37 and 60 °C. The employed on-resonance $R_{1\rho}$ and ^{1}H CEST pulse sequences have been modified from Schlagnitweit et al.[40] CEST profiles and pseudo-2D ^{1}H R_1 and pseudo-3D on-resonance $R_{1\rho}$ experiments were performed, processed, and analyzed as previously described.[20]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsphyschemau.1c00050.

^{1}H−^{1}H, ^{1}H−^{13}C, and ^{1}H−^{15}N 2D correlation spectra, schematic comparison of ^{1}H, ^{13}C, and ^{15}N chemical shift perturbations, concentration-dependent UV/vis melting analysis, temperature-dependent CEST profiles, and van't Hoff plots for all available sites, on-resonance $R_{1\rho}$ relaxation dispersion profiles for all samples, CEST-derived dsDNA and ssDNA chemical shifts and relaxation parameters (R_1 and R_2), and CEST-derived thermodynamic and kinetic parameters for all samples (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Petra Rovó − *Faculty of Chemistry and Pharmacy, Department of Chemistry, Ludwig-Maximilians-Universität München, 81377 Munich, Germany; Center for Nanoscience (CeNS), Faculty of Physics, Ludwig-Maximilians-Universität München, 80799 Munich, Germany; Institute of Science and Technology Austria (ISTA), 3400 Klosterneuburg, Austria;* orcid.org/0000-0001-8729-7326; Email: petra.rovo@ist.ac.at

Authors

Romeo C. A. Dubini − *Faculty of Chemistry and Pharmacy, Department of Chemistry, Ludwig-Maximilians-Universität München, 81377 Munich, Germany; Center for Nanoscience (CeNS), Faculty of Physics, Ludwig-Maximilians-Universität München, 80799 Munich, Germany;* orcid.org/0000-0001-6045-271X

Eva Korytiaková − *Faculty of Chemistry and Pharmacy, Department of Chemistry, Ludwig-Maximilians-Universität München, 81377 Munich, Germany*

Thea Schinkel − *Faculty of Chemistry and Pharmacy, Department of Chemistry, Ludwig-Maximilians-Universität München, 81377 Munich, Germany;* orcid.org/0000-0002-1891-2660

Pia Heinrichs − *Faculty of Chemistry and Pharmacy, Department of Chemistry, Ludwig-Maximilians-Universität München, 81377 Munich, Germany;* orcid.org/0000-0002-3128-6676

Thomas Carell − *Faculty of Chemistry and Pharmacy, Department of Chemistry, Ludwig-Maximilians-Universität München, 81377 Munich, Germany*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsphyschemau.1c00050

Author Contributions

[§]T.S. and P.H. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Markus Müller for valued discussions and Felix Xu for assistance in the measurement of UV/vis melting profiles. This work was supported in part by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) − SFB 1309-325871075, EU-ITN LightDyNAmics (ID: 765266), the ERC-AG EpiR (ID: 741912), the Center for NanoScience, the Excellence Clusters CIPSM, and the Fonds der Chemischen Industrie. Open access funding provided by Institute of Science and Technology Austria (ISTA).

■ REFERENCES

(1) Kumar, S.; Chinnusamy, V.; Mohapatra, T. Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond. *Front. Genet.* **2018**, *9*, 1−14.

(2) Bilyard, M. K.; Becker, S.; Balasubramanian, S. Natural, modified DNA bases. *Curr. Opin. Chem. Biol.* **2020**, *57*, 1−7.

(3) Chen, Y.; Hong, T.; Wang, S.; Mo, J.; Tian, T.; Zhou, X. Epigenetic modification of nucleic acids: from basic studies to medical applications. *Chem. Soc. Rev.* **2017**, *46*, 2844−2872.

(4) Carell, T.; Kurz, M. Q.; Müller, M.; Rossa, M.; Spada, F. Non-canonical Bases in the Genome: The Regulatory Information Layer in DNA Angewandte. *Angew. Chem., Int. Ed.* **2018**, *57*, 4296−4312.

(5) Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science* **2011**, *333*, 1300−1304.

(6) Wu, X.; Zhang, Y. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.* **2017**, *18*, 517−534.

(7) Bachman, M.; Uribe-Lewis, S.; Yang, X.; Burgess, H. E.; Iurlaro, M.; Reik, W.; Murrell, A.; Balasubramanian, S. 5-Formylcytosine can be a stable DNA modification in mammals. *Nat. Chem. Biol.* **2015**, *11*, 555−557.

(8) Su, M.; Kirchner, A.; Stazzoni, S.; Müller, M.; Wagner, M.; Schröder, A.; Carell, T. 5-Formylcytosine Could Be a Semipermanent Base in Specific Genome Sites. *Angew. Chem., Int. Ed.* **2016**, *55*, 11797−11800.

(9) Koivunen, P.; Laukka, T. The TET enzymes. *Cell. Mol. Life Sci.* **2018**, *75*, 1339−1348.

(10) Guo, M.; Li, X.; Zhang, L.; Liu, D.; Du, W.; Yin, D.; Lyu, N.; Zhao, G.; Guo, C.; Tang, D. Accurate quantification of 5-

Methylcytosine, 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxylcytosine in genomic DNA from breast cancer by chemical derivatization coupled with ultra performance liquid chromatography-electrospray quadrupole time. *Oncotarget* **2017**, *8*, 91248−91257.
(11) Eleftheriou, M.; Pascual, A. J.; Wheldon, L. M.; Perry, C.; Abakir, A.; Arora, A.; Johnson, A. D.; Auer, D. T.; Ellis, I. O.; Madhusudan, S.; et al. 5-Carboxylcytosine levels are elevated in human breast cancers and gliomas. *Clin. Epigenet.* **2015**, *7*, 88.
(12) Hashimoto, H.; Olanrewaju, Y. O.; Zheng, Y.; Wilson, G. G.; Zhang, X.; Cheng, X. Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence. *Genes Dev.* **2014**, *28*, 2304−2313.
(13) Wang, D.; Hashimoto, H.; Zhang, X.; Barwick, B. G.; Lonial, S.; Boise, L. H.; Vertino, P. M.; Cheng, X. MAX is an epigenetic sensor of 5-carboxylcytosine and is altered in multiple myeloma. *Nucleic Acids Res.* **2017**, *45*, 2396−2407.
(14) Wheldon, L. M.; Abakir, A.; Ferjentsik, Z.; Dudnakova, T.; Strohbuecker, S.; Christie, D.; Dai, N.; Guan, S.; Foster, J. M.; Corrêa, I. R.; et al. Transient accumulation of 5-carboxylcytosine indicates involvement of active demethylation in lineage specification of neural stem cells. *Cell Rep* **2014**, *7*, 1353−1361.
(15) Kellinger, M. W.; Song, C. X.; Chong, J.; Lu, X. Y.; He, C.; Wang, D. 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **2012**, *19*, 831−833.
(16) Song, J.; Pfeifer, G. P. Are there specific readers of oxidized 5-methylcytosine bases? *BioEssays* **2016**, *38*, 1038−1047.
(17) Raiber, E. A.; Murat, P.; Chirgadze, D. Y.; Beraldi, D.; Luisi, B. F.; Balasubramanian, S. 5-formylcytosine alters the structure of the DNA double helix. *Nat. Struct Mol. Biol.* **2015**, *22*, 44−49.
(18) Wang, S.; Long, Y.; Wang, J.; Ge, Y.; Guo, P.; Liu, Y.; Tian, T.; Zhou, X. Systematic Investigations of Different Cytosine Modifications on CpG Dinucleotide Sequences: The Effects on the B-Z. Transition. *J. Am. Chem. Soc.* **2014**, *136*, 56.
(19) Hardwick, J. S.; Ptchelkine, D.; El-Sagheer, A. H.; Tear, I.; Singleton, D.; Phillips, S. E.; Lane, A. N.; Brown, T. 5-Formylcytosine does not change the global structure of DNA. *Nat. Struct. Mol. Biol.* **2017**, *24*, 544−552.
(20) Dubini, R. C.; Schön, A.; Müller, M.; Carell, T.; Rovó, P. Impact of 5-formylcytosine on the melting kinetics of DNA by 1H NMR chemical exchange. *Nucleic Acids Res.* **2020**, *48*, 8796−8807.
(21) Sanstead, P. J.; Ashwood, B.; Dai, Q.; He, C.; Tokmakoff, A. Oxidized Derivatives of 5-Methylcytosine Alter the Stability and Dehybridization Dynamics of Duplex DNA. *J. Phys. Chem. B* **2020**, *124*, 1160−1174.
(22) Fu, T.; Liu, L.; Yang, Q.-l.; Wang, Y.; Xu, P.; Zhang, L.; Liu, S.; Dai, Q.; Ji, Q.; Xu, G.-l.; et al. Thymine DNA glycosylase recognizes the geometry alteration of minor grooves induced by 5-formylcytosine and 5-carboxylcytosine. *Chem. Sci.* **2019**, *10*, 7407−7417.
(23) Szulik, M. W.; Pallan, P. S.; Nocek, B.; Voehler, M.; Banerjee, S.; Brooks, S.; Joachimiak, A.; Egli, M.; Eichman, B. F.; Stone, M. P. Differential Stabilities and Sequence-Dependent Base Pair Opening Dynamics of Watson-Crick Base Pairs with 5-Hydroxymethylcytosine, 5-Formylcytosine, or 5-Carboxylcytosine. *Biochemistry* **2015**, *54*, 1294.
(24) Sumino, M.; Ohkubo, A.; Taguchi, H.; Seio, K.; Sekine, M. Synthesis and properties of oligodeoxynucleotides containing 5-carboxy-2−deoxycytidines. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 274−277.
(25) Dai, Q.; Sanstead, P. J.; Peng, C. S.; Han, D.; He, C.; Tokmakoff, A. Weakened N3 Hydrogen Bonding by 5-Formylcytosine and 5-Carboxylcytosine Reduces Their Base-Pairing Stability. *ACS Chem. Biol.* **2016**, *11*, 470−477.
(26) Maiti, A.; Michelson, A. Z.; Armwood, C. J.; Lee, J. K.; Drohat, A. C. Divergent mechanisms for enzymatic excision of 5-formylcytosine and 5-carboxylcytosine from DNA. *J. Am. Chem. Soc.* **2013**, *135*, 15813−15822.
(27) Ashwood, B.; Sanstead, P. J.; Dai, Q.; He, C.; Tokmakoff, A. 5-Carboxylcytosine and Cytosine Protonation Distinctly Alter the Stability and Dehybridization Dynamics of the DNA Duplex. *J. Phys. Chem. B* **2020**, *124*, 627−640.
(28) Shibutani, T.; Ito, S.; Toda, M.; Kanao, R.; Collins, L. B.; Shibata, M.; Urabe, M.; Koseki, H.; Masuda, Y.; Swenberg, J. A.; et al. Guanine-5-carboxylcytosine base pairs mimic mismatches during DNA replication. *Sci. Rep.* **2015**, *4*, 5220.
(29) Vallurupalli, P.; Bouvignies, G.; Kay, L. E. Studying "Invisible" Excited Protein States in Slow Exchange with a Major State Conformation. *J. Am. Chem. Soc.* **2012**, *134*, 8148−8161.
(30) Liu, B.; Rangadurai, A.; Shi, H.; Al-Hashimi, H. M. Rapid assessment of Watson−Crick to Hoogsteen exchange in unlabeled DNA duplexes using high-power SELOPE imino ^{1}H CEST. *Magn. Reson.* **2021**, *2*, 715−731.
(31) Nikolova, E. N.; Kim, E.; Wise, A. A.; O'Brien, P. J.; Andricioaei, I.; Al-Hashimi, H. M. Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* **2011**, *470*, 498−502.
(32) Zhou, H.; Sathyamoorthy, B.; Stelling, A.; Xu, Y.; Xue, Y.; Pigli, Y. Z.; Case, D. A.; Rice, P. A.; Al-Hashimi, H. M. Characterizing Watson-Crick versus Hoogsteen Base Pairing in a DNA-Protein Complex Using Nuclear Magnetic Resonance and Site-Specifically 13C- and 15N-Labeled DNA. *Biochemistry* **2019**, *58*, 1963−1974.
(33) Galindo-Murillo, R.; Roe, D. R.; Cheatham, T. E., III On the absence of intrahelical DNA dynamics on the μs to ms timescale. *Nat. Commun.* **2014**, *5*, 5152.
(34) Choi, S.-R.; Kim, N.-H.; Jin, H.-S.; Seo, Y.-J.; Lee, J.; Lee, J.-H. Base-pair Opening Dynamics of Nucleic Acids in Relation to Their Biological Function. *Comput. Struct. Biotechnol. J.* **2019**, *17*, 797−804.
(35) Schanda, P.; Brutscher, B. Very Fast Two-Dimensional NMR Spectroscopy for Real-Time Investigation of Dynamic Events in Proteins on the Time Scale of Seconds. *J. Am. Chem. Soc.* **2005**, *127*, 8014−8015.
(36) Zarycz, M. N. C.; Fonseca Guerra, C. NMR ^{1}H-Shielding Constants of Hydrogen-Bond Donor Reflect Manifestation of the Pauli Principle. *J. Phys. Chem. Lett.* **2018**, *9*, 3720−3724.
(37) Marušič, M.; Schlagnitweit, J.; Petzold, K. RNA Dynamics by NMR Spectroscopy. *ChemBioChem.* **2019**, *20*, 2685−2710.
(38) Shi, H.; Liu, B.; Nussbaumer, F.; Rangadurai, A.; Kreutz, C.; Al-Hashimi, H. M. NMR Chemical Exchange Measurements Reveal That N^{6}-Methyladenosine Slows RNA Annealing. *J. Am. Chem. Soc.* **2019**, *141*, 19988−19993.
(39) Wang, Y.; Han, G.; Jiang, X.; Yuwen, T.; Xue, Y. Chemical shift prediction of RNA imino groups: application toward characterizing RNA excited states. *Nat. Commun.* **2021**, *12*, 1595.
(40) Schlagnitweit, J.; Steiner, E.; Karlsson, H.; Petzold, K. Efficient Detection of Structure and Dynamics in Unlabeled RNAs: The SELOPE Approach. *Chem.—Eur. J.* **2018**, *24*, 6067−6070.
(41) Hardwick, J. S.; Lane, A. N.; Brown, T. Epigenetic Modifications of Cytosine: Biophysical Properties, Regulation, and Function in Mammalian DNA. *BioEssays* **2018**, *40*, 1700199.
(42) Zhu, Q.; Stöger, R.; Alberio, R. A Lexicon of DNA Modifications: Their Roles in Embryo Development and the Germline. *Front. Cell Dev. Biol.* **2018**, *6*, 24.
(43) Wu, H.; Wu, X.; Shen, L.; Zhang, Y. Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing. *Nat. Biotechnol.* **2014**, *32*, 1231−1240.
(44) Zhu, C.; Gao, Y.; Guo, H.; Xia, B.; Song, J.; Wu, X.; Zeng, H.; Kee, K.; Tang, F.; Yi, C. Single-Cell 5-Formylcytosine Landscapes of Mammalian Early Embryos and ESCs at Single-Base Resolution. *Cell Stem Cell* **2017**, *20*, 720−731.
(45) Lewis, L. C.; Lo, P. C. K.; Foster, J. M.; Dai, N.; Corrêa, I. R.; Durczak, P. M.; Duncan, G.; Ramsawhook, A.; Aithal, G. P.; Denning, C.; et al. Dynamics of 5-carboxylcytosine during hepatic differentiation: Potential general role for active demethylation by DNA repair in lineage specification. *Epigenetics* **2017**, *12*, 277−286.
(46) Ngo, T. T.; Yoo, J.; Dai, Q.; Zhang, Q.; He, C.; Aksimentiev, A.; Ha, T. Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nat. Commun.* **2016**, *7*, 10813.

https://doi.org/10.1021/acsphyschemau.1c00050
ACS Phys. Chem Au XXXX, XXX, XXX−XXX

(47) Lee, S. H.; Griffiths, J. R. How and Why Are Cancers Acidic? Carbonic Anhydrase IX and the Homeostatic Control of Tumour Extracellular pH. *Cancers (Basel)* **2020**, *12*, 1616.
(48) Swietach, P.; Vaughan-Jones, R. D.; Harris, A. L.; Hulikova, A. The chemistry, physiology and pathology of pH in cancer. *Philos. Trans. R. Soc. B Biol. Sci.* **2014**, *369*, 20130099.
(49) Spruijt, C. G.; Gnerlich, F.; Smits, A. H.; Pfaffeneder, T.; Jansen, P. W.; Bauer, C.; Münzel, M.; Wagner, M.; Müller, M.; Khan, F.; et al. Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives. *Cell* **2013**, *152*, 1146−1159.
(50) Dodd, T.; Yan, C.; Kossmann, B. R.; Martin, K.; Ivanov, I. Uncovering universal rules governing the selectivity of the archetypal DNA glycosylase TDG. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 5974−5979.
(51) Coey, C. T.; Malik, S. S.; Pidugu, L. S.; Varney, K. M.; Pozharski, E.; Drohat, A. C. Structural basis of damage recognition by thymine DNA glycosylase: Key roles for N-terminal residues. *Nucleic Acids Res.* **2016**, *44*, 10248−10258.
(52) Kanaan, N.; Imhof, P. Interactions of the DNA Repair Enzyme Human Thymine DNA Glycosylase with Cognate and Noncognate DNA. *Biochemistry* **2018**, *57*, 5654−5665.
(53) Schröder, A. S.; Steinbacher, J.; Steigenberger, B.; Gnerlich, F. A.; Schiesser, S.; Pfaffeneder, T.; Carell, T. Synthesis of a DNA Promoter Segment Containing All Four Epigenetic Nucleosides: 5-Methyl-, 5-Hydroxymethyl-, 5-Formyl-, and 5-Carboxy- 2'-Deoxycytidine. *Angew. Chem., Int. Ed.* **2014**, *53*, 315−318.
(54) Favier, A.; Brutscher, B. NMRlib: user-friendly pulse sequence tools for Bruker NMR spectrometers. *J. Biomol. NMR* **2019**, *73*, 199−211.

https://doi.org/10.1021/acsphyschemau.1c00050
ACS Phys. Chem Au XXXX, XXX, XXX−XXX

3.7 Synthesis and Structure Elucidation of the Human tRNA Nucleoside Mannosyl-Queuosine

M. Hillmeier, M. Wagner, T. Ensfelder, E. Korytiaková, P. Thumbs, M. Müller, T. Carell, *Nat. Commun.* **2021**, *12*, 7123.

Prolog

In RNA exisitiert neben den vier kanonischen Basen, C, G, A und U eine Vielzahl weiterer chemisch modifizierter Nukleoside. Diese posttranskriptionel eingeführten Modifikationen geraten immer mehr in das Interesse der Forschung, insbesondere da ein Fehlen vieler dieser Modifikationen zunehmend in Zusammenhang mit unterschiedlichen Erkrankungen gebracht werden kann. Von den verschiedenen in der Zelle auftretenden RNA Spezies ist tRNA eine der am stärksten modifizierten RNAs. tRNA oder transfer-RNA ist im Rahmen der Proteinbiosynthese für die korrekte Codon Erkennung und den Einbau der zugehörigen Aminosäure in das entstehende Protein erforderlich. Hierzu sind auch Modifikationen vor allem im Anticodon Bereich erforderlich. Queuosin ist eine hypermodifizierte RNA Base, welche an Position 34 und damit im Anitcodon von tRNA-Aps, -His, Asn- und Tyr in verschiedensten eukaryotischen und prokaryotischen Spezies, einschließlich des Menschen, vorliegt. Zusätzlich zu Queuosin selbst, exisieren zwei weitere Hexose-modifizierte Derivate, Mannosyl- und Galactosyl-Queuosin. Im Rahmen dieser Arbeit wurde die bisher bekannte Struktur von Mannosyl-Queuosin überprüft.[157] Hierzu wurden verschiedene mögliche Regio- und Stereosiomere des Mannosyl-Queuosins totalsynthetisch dargestellt und anschließend mittels UHPLC-MS/MS Koinjektionsexperimenten mit dem natürlich vorkommenden Nukleosid verglichen. Dabei stellte sich heraus, dass die bisher korrekte Struktur von Mannosyl-Queuosin eine α-Allyl Konnektivität aufweist. Dieses Ergebnis steht im Kontrast zur bisher angenommenen Struktur und widerlegt diese eindeutig. Die in dieser Studie aufgelöste Struktur von Mannosyl-Queuosin stellen einen wichtigen Meilenstein zur weiteren Erforschung der Funktion dieser hypermodifizierten Nukleoside dar.

Autorenbeitrag

In dieser Arbeit war ich für die Entwicklung der massenspektrometrischen Methoden mittels UHPLC-MS/MS für Mannosyl-Queuosin-Derivate sowie die Durchführung der Koinjektions-experimente der verdauten RNA und der synthetisierten Standards zuständig. Der Haupteil der Arbeit dieser Studie, die Entwicklung und Durchführung der Synthese der Mannosyl-Queuosin-Derivate erfolgte durch *Markus Hillmeier*. *Mirko Wagner* hat massen-

spektrometrische LC-MS Koinjektionsexperimente durchgeführt. *Timm Ensfelder* hat zellbiologische Verfütterungsexperimente durchgeführt. *Peter Thumbs* hat die Synthese von Galactosyl-Queuosin entwickelt.

Lizenz

Kopie der Publikation mit Erlaubnis des Verlags; Copyright 2021 *Nature Publishing Group*.

DOI: 10.1038/s41467-021-27371

Artikel einsehbar unter:

https://www.nature.com/articles/s41467-021-27371-9

Die *Supporting Information* dieser Veröffentlichung ist einsehbar unter:

https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-021-27371-9/MediaObjects/41467_2021_27371_MOESM1_ESM.pdf

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-021-27371-9 OPEN

Synthesis and structure elucidation of the human tRNA nucleoside mannosyl-queuosine

Markus Hillmeier[1], Mirko Wagner[1], Timm Ensfelder[1], Eva Korytiakova[1], Peter Thumbs[1], Markus Müller[1] & Thomas Carell[1✉]

Queuosine (Q) is a structurally complex, non-canonical RNA nucleoside. It is present in many eukaryotic and bacterial species, where it is part of the anticodon loop of certain tRNAs. In higher vertebrates, including humans, two further modified queuosine-derivatives exist - galactosyl- (galQ) and mannosyl-queuosine (manQ). The function of these low abundant hypermodified RNA nucleosides remains unknown. While the structure of galQ was elucidated and confirmed by total synthesis, the reported structure of manQ still awaits confirmation. By combining total synthesis and LC-MS-co-injection experiments, together with a metabolic feeding study of labelled hexoses, we show here that the natural compound manQ isolated from mouse liver deviates from the literature-reported structure. Our data show that manQ features an α-allyl connectivity of its sugar moiety. The yet unidentified glycosylases that attach galactose and mannose to the Q-base therefore have a maximally different constitutional connectivity preference. Knowing the correct structure of manQ will now pave the way towards further elucidation of its biological function.

[1] Department Chemie, Ludwig-Maximilians-Universität München, Butenandtstraße 5-13, 81377 München, Germany. ✉email: thomas.carell@cup.uni-muenchen.de

Ribonucleic acid (RNA) is a central molecule of life, linking the genotype to the phenotype by integrating both catalytic and coding properties in the synthesis of proteins[1,2]. To fulfil the plethora of functions known for RNA today, a huge chemical diversity has developed regarding the nucleobases through the evolution of life[3,4]. The most densely modified RNA molecules are the tRNAs[5]. Those small adaptor molecules feed the required amino acids to the growing peptide chain in the ribosome. Of particular interest are the non-canonical nucleosides that are found in the anticodon loop of tRNAs, in particular at position 34, known as the Wobble position of the anticodon[6]. Here, any chemical modification has a direct impact on the coding potential of the tRNA[6–9]. Queuosine (1, Q, Fig. 1a) is a particularly complex non-canonical nucleoside[10–12]. It is found in many bacterial and eukaryotic species, where it is located at position 34 of the anticodon loop of $tRNA^{Tyr}$, $tRNA^{Asp}$, $tRNA^{His}$ and $tRNA^{Asn}$[13,14].

In eukaryotes, queuosine 1 is found in both cytosolic and mitochondrial tRNA[15]. It is biochemically derived from guanosine (G) and characterised by the exchange of the N7-nitrogen for a C7-carbon atom and the addition of a 1(S)-amino-2(R),3(S)-dihydroxycyclopent-4-ene unit to the C7-position via a methylene linker[12,16]. The presence of the Q-base in position 34 of the anticodon equips the corresponding tRNAs with the possibility to decode synonymous codons by wobble base pairing[17,18]. The Q-base is furthermore affecting translational speed[19], decoding fidelity[20], and it is essential for survival particularly in the absence of sufficient tyrosine[21,22]. Eukaryotes are unable to biosynthesize Q, which forces them to acquire it from prokaryotic sources[21,23–25]. This establishes a link between the central process of translation and the gut microbiome[26]. Defective RNA modification processes are increasingly recognised as drivers for severe diseases[27,28].

In higher vertebrates, including humans, two glycosylated queuosine derivatives exist, which have yet unidentified functions. The first of these hypermodified Q derivatives is found in cytosolic $tRNA^{Tyr}$, where the Q-base is modified with a galactose residue (galQ, 2, Fig. 1b)[10,29]. The cytosolic $tRNA^{Asp}$, in contrast, contains a Q-derivative, which is suggested to be mannosylated (manQ, Fig. 1c, d). A first effort to elucidate the structure of these nucleosides was reported in 1976 by Kasai et al. They derived the structure of galQ and proposed a structure for manQ based on NMR- and MS-data that was obtained from isolated material from rabbit liver. This data lead to the proposal of structure 3 for natural manQ[29]. Here, we report the total synthesis of the human natural product manQ and show that the originally proposed structure 3 (Fig. 1c) needs to be revised in two important aspects: in contrast to the original structure suggestion, we show that the mannose glycosidic bond is not β- but α-configured. Furthermore, and in contrast to β-galQ, the mannose is attached not to the homoallylic, but to the allylic hydroxyl group at C3. manQ has consequently the structure 4 shown in Fig. 1d. This result reveals that the yet unknown galactosyl and mannosyl transferases that attach the respective hexose to the Q-nucleobase in $tRNA^{Tyr}$ and $tRNA^{Asp}$, are able to differentiate the two hydroxyl groups (allyl vs. homoallyl) at the cyclopentene substructure for yet unknown reasons.

Results and discussion

Synthesis of literature-reported manQ. The compounds galQ and manQ are found in the anticodon loop of one cytosolic tRNA each[30]. Consequently, the amounts that can be isolated from nature (e.g. porcine liver) are very small, so that a structure elucidation from isolated material only is very difficult. $tRNA^{Asp}$, which contains one molecule of manQ per tRNA, can be isolated[30], but in our hands it was impossible to obtain sufficient amounts for a full structure elucidation. Therefore, synthesis of the proposed manQ compound and comparison of the synthetic standard with natural material by LC-MS was the method of choice for us to investigate the manQ structure. The synthesis of compound 3, proposed to be natural manQ, was performed as depicted in Fig. 2a.

For the synthesis of the difficult-to-access β-mannosyl connectivity, we employed the Crich-method[31], which required the preparation of the mannosyl donor 5. This was achieved along the

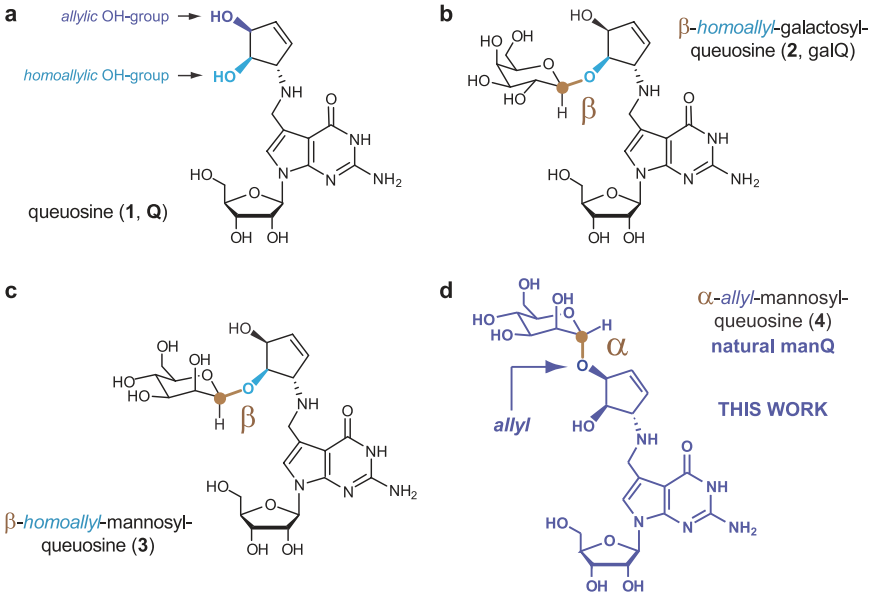


Fig. 1 Nucleosides of the queuosine family present in higher vertebrates. a Queuosine 1 is derived from a 7-deazaguanosine base with a 1(S)-amino-2(R),3(S)-dihydroxycyclopent-4-ene unit attached via a methylene linker to the C7-position of the 7-deazaguanosine. Possible glycosylation sites are depicted in dark blue (allylic hydroxyl group) and light blue (homoallylic hydroxyl group); **b** Structure of the naturally occurring nucleoside β-homoallyl-galQ 2; **c** Structure 3 shows the originally proposed structure of manQ, which we revise in this work; **d** Corrected structure of natural manQ 4 featuring an α-allylic connectivity.

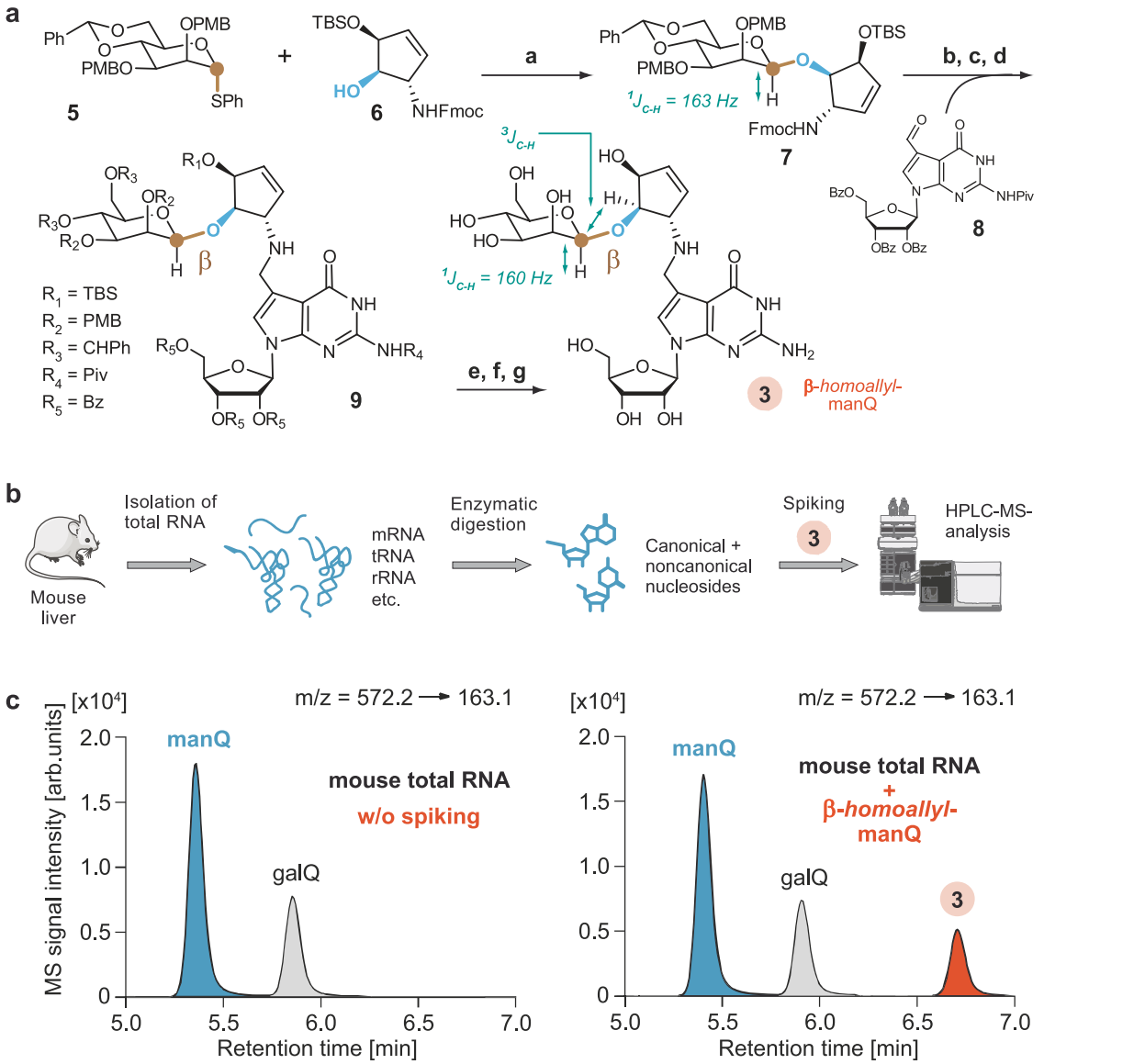


Fig. 2 Attempt to confirm the originally proposed structure of manQ. a Synthetic route towards β-homoallyl-manQ featuring a β-selective mannosylation key step; a: AgOTf, PhSCl, DTBMP, DCM, −78 °C, 2 h, 62%; b: $HNMe_2$, THF, rt, 1 h; c: 8, MeOH, rt, 5 h; d: $NaBH_4$, MeOH, 0 °C, 1 h; e: NaOMe, MeOH, rt, 5 h; f: TFA, DCM, 0 °C, 20 min; g: HF · pyridine, pyridine, rt, 18 h, 22% (5 steps); **b** Workflow for the comparison of the synthetic β-homoallyl-manQ standard with natural manQ via isolation of total RNA from mouse liver, digestion to the nucleoside level and LC-MS-co-injection experiments; **c** Chromatograms resulting from the co-injection experiments analysed by UHPLC-MS/MS; left: digested total RNA from mouse liver shows two peaks, natural manQ and galQ,; right: digested total RNA from mouse liver spiked with the synthetic standard β-homoallyl-manQ 3 shows the appearance of an additional peak, thereby disproving the originally proposed structure for manQ. m/z = 572.2 is both the mass of the protonated manQ or galQ nucleoside, while m/z = 163.1 is the mass of a specific fragment ion formed during the fragmentation of these molecules in the mass spectrometer (see Supplementary Fig. 1). This mass transition was used by us for manQ and galQ detection via MS/MS.

literature-reported synthesis[31,32]. The mannosylation reaction of the Fmoc- and TBS- protected 1(S)-amino-2(R),3(S)-dihydroxycyclopent-4-ene unit 6 with the mannosyl donor 5 required an activation with Ag-triflate in the presence of phenylthiochloride and 2,6-di-tert-butyl-4-methylpyridine (DTBMP) at −78 °C in dry dichloromethane. The reaction gave the mannoside 7 in 62% yield, and the β-configuration of its anomeric centre was confirmed by NMR-spectroscopy based on the coupling constant of $^1J_{C1\text{-}H1} = 163$ Hz, which is typical for β-mannosides[33]. We next cleaved the Fmoc group of the mannoside 7 with dimethylamine (10% DMA in THF, rt, 1 h) and performed the reductive amination with the benzoyl- and pivaloyl-protected 7-formyl-7-deazaguanosine compound 8 (MeOH, rt, 5 h, then $NaBH_4$, 0 °C, 1 h) that was prepared as recently described by us[10]. This afforded the fully protected manQ compound 9. Cleavage of the benzoyl-protecting groups at the ribose was performed under Zemplén-conditions (0.5 M NaOMe, MeOH, rt, 5 h). The PMB-ethers were deprotected with trifluoro acetic acid in dichloromethane (10%, 0 °C, 20 min). Finally, we removed the TBS-protecting groups with HF in pyridine (rt, 18 h). This furnished the final manQ compound 3 with mannose being β-configured and attached to the homoallylic hydroxyl group. As expected, compound 3 has a $^1J_{C\text{-}H}$-coupling constant of 160 Hz at the anomeric centre, thereby proving its β-configuration. Furthermore, the observed $^3J_{C\text{-}H}$-HMBC coupling in compound 3 between the mannosyl C1-carbon and the homoallylic hydrogen C4H of the cyclopentene

ring (see Supplementary Information 4) unambiguously proved that the mannosyl residue is indeed connected to the homoallylic hydroxyl group.

In order to compare the synthetic manQ compound 3 with the natural material, we performed an LC-MS co-injection study. To this end, we isolated the total RNA from mouse liver and enzymatically digested this RNA down to the individual nucleosides as recently described by us[10]. This procedure afforded a mixture of all nucleosides present in the total RNA pool. Subsequent analysis of the resulting nucleoside mixture was performed via UHPLC-MS/MS with a triple quadrupole mass spectrometer (QQQ) set to monitor a specific molecular fragmentation reaction of the glycosylated Q-derivatives. Using collision-induced dissociation, these nucleosides undergo two heterolytic bond dissociations, leading to a loss of both the ribose sugar and the glycosylated cyclopentene unit. Monitoring the corresponding mass transition of m/z = 572.2 → 163.1 allows a sensitive detection of these compounds (see Fig. 2b and Supplementary Fig. 1). With an UHPLC separation gradient of 0 → 2% (v/v) MeCN/H_2O in 0 → 8 min on a Poroshell 120 SB-C8 column, all hexose-modified queuosine derivatives can be separated chromatographically (see Fig. 2c). Indeed, when analysing the mouse liver sample, we detected two clearly separated signals in the RNA nucleoside pool with galQ 2 eluting at 5.9 min, and the natural manQ compound eluting at 5.4 min, respectively. The compound eluting at 5.9 min was unambiguously identified as galQ 2 with the help of a synthetic reference compound (see Supplementary Fig. 2). Upon co-injection of the synthetic material 3, we were expecting to see again two signals with the manQ signal having gained in intensity. To our surprise, however, we noted that this is not the case. Instead, the co-injection experiment provided a third signal with a retention time of 6.8 min, clearly separated from both natural manQ and galQ 2. This result unequivocally shows that manQ 3 is NOT identical to the natural manQ compound. Therefore, the structure proposal for manQ reported in literature must be incorrect.

Puzzled by this result, we initially reasoned that the compound originally identified as manQ may contain a different sugar than mannose, even though a previous study showed that a mannosyl moiety can enzymatically be transferred to the Q-base from GDP-mannose[34]. To verify that manQ contains indeed a mannose sugar, we performed a metabolic labelling study (Fig. 3a): We added different isotope-labelled sugars to a HEK293T cell culture and subsequently isolated the total RNA of the cells to see if the natural manQ had incorporated the isotope labels. During these experiments, the cell culture medium was additionally supplemented with high concentrations of unlabelled glucose or mannose as carbon source in order to suppress the metabolic conversion of the labelled sugar into other carbohydrates[35]. Without these metabolic suppressors, we observed in our experiments an unwanted isotope scrambling process that jeopardised the experiments. The results of our study are depicted in Fig. 3b. It is clearly evident that feeding of $^{13}C_6$-galactose and $^{13}C_6$-glucose in the presence of the metabolic suppressors glucose or mannose, respectively, gave little or no incorporation of ^{13}C into the isolated natural manQ compound. In contrast, feeding of $^{13}C_6$-mannose in combination with the metabolic suppressor glucose quickly led to the time-dependent formation of $^{13}C_6$-manQ, thereby confirming that the sugar connected to the Q-base in manQ is indeed mannose.

We therefore reasoned that natural manQ differs from the reported structure 3 likely regarding the configuration of the anomeric centre and/or the connectivity at the cyclopentene moiety (allyl versus homoallyl mannoside). With 3 proven not to be the natural compound, we were left with the three remaining structures 4 (α-allyl-connectivity), 10 (β-allyl-connectivity) and

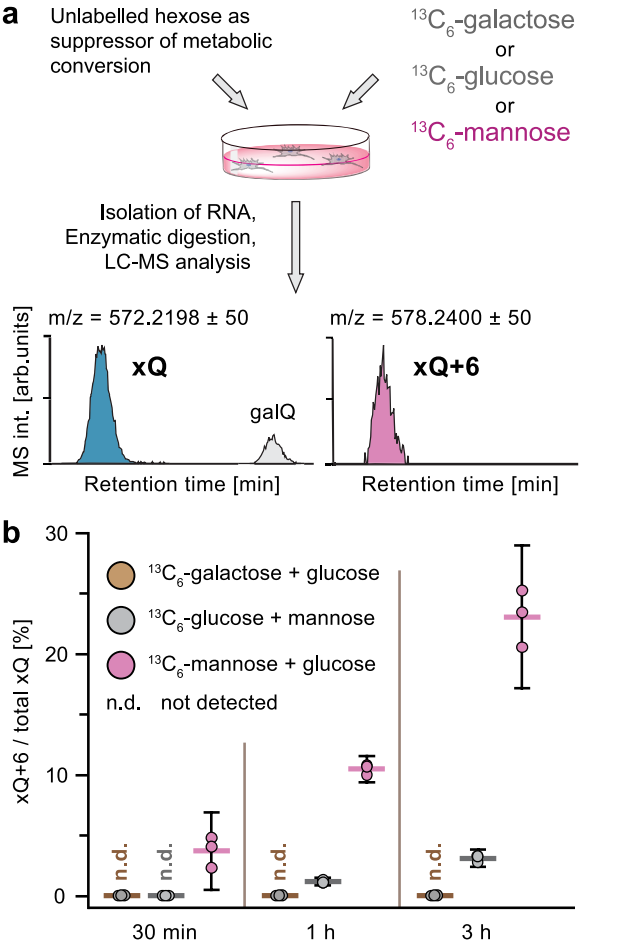


Fig. 3 Confirmation of the incorporated hexose sugar as being mannose. **a** Workflow of the metabolic labelling study providing $^{13}C_6$-galactose, $^{13}C_6$-glucose, or $^{13}C_6$-mannose in combination with an unlabelled carbon source (glucose or mannose) that suppresses metabolic interconversion of the carbohydrates and hence isotope scrambling; **b** Results of the LC-MS experiments showing quick incorporation of heavy sugar when feeding $^{13}C_6$-mannose in presence of glucose, thereby confirming the identity of xQ as manQ; three independent experimental replicates were performed for each condition tested (coloured dots). Mean values are indicated by coloured horizontal lines, while whiskers enclose the 95%-confidence interval.

11 (α-homoallyl-connectivity, see Fig. 4a). For their synthesis, it was necessary to first develop syntheses for the properly protected 1-amino-2,3-dihydroxycyclopent-4-enes 13–15 (Fig. 4b). To access these compounds, we started with the Fmoc-protected 1-amino-2,3-dihydroxypent-4-ene 12. For the synthesis of the 1-Fmoc-3-PMB-protected cyclopent-4-ene 13, needed for the synthesis of 11, we first protected in 12 both OH groups as an anisaldehyde acetal (anisaldehyde dimethyl acetal, CSA, rt, 2 h, 98%), followed by a selective reductive opening of the homoallylic OH group with DIBAL-H. This gave the 1-Fmoc-3-PMB-protected-dihydroxycyclopent-4-ene 13 (DCM, −78 °C, 3 h, 85%). For 14, we used 13 as the starting material. Protection of the homoallylic hydroxyl group in 13 with SEM-Cl (NBu_4I, pyridine, DMF, 70 °C, 18 h, 75%), followed by cleavage of the p-methoxybenzylether (10% TFA in DCM, 0 °C, 5 min, 83%) furnished compound 14. Reaction of 12 with TBSOTf at −55 °C provided after 15 min predominantly the 1-Fmoc-3-TBS-protected compound 6 as the kinetic product, which we had used before for the synthesis of the literature-reported structure of

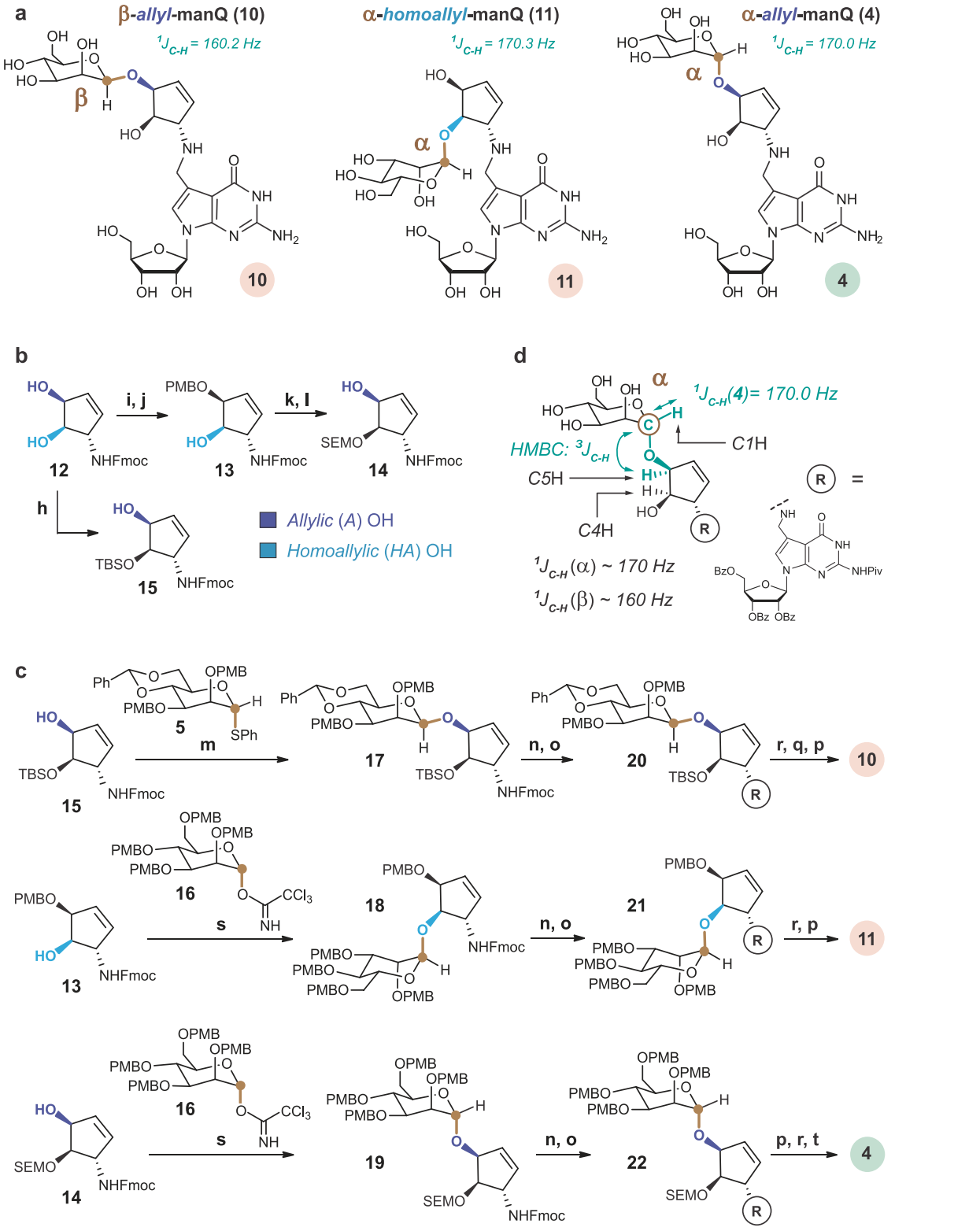


Fig. 4 Alternative manQ structures and their syntheses. a manQ target compounds: β-allyl-manQ 10, α-homoallyl-manQ 11 and the correct structure 4 of natural manQ featuring an α-allyl-connectivity; **b** Regioselective protection of cyclopentene 12 for the synthesis of precursors 13–15; h: TBSOTf, DMF, −10 °C, 2 h, 41%; i: anisaldehyde dimethyl acetal, CSA, DMF, rt, 2 h, 98%; j: DIBAL-H, DCM, −78 °C, 3 h, 85%; k: SEM-Cl, NBu_4I, pyridine, DMF, 70 °C, 18 h, 75%; l: TFA, DCM, 0 °C, 5 min, 83%; **c** Synthesis of the manQ compounds 10, 11 and 4 from 13, 14 and 15 via an appropriate stereoselective glycosylation step and subsequent reductive amination; m: AgOTf, PhSCl, DTBMP, DCM, −78 °C, 2 h, 52%; n: DBU, MeCN, rt, 1 h (10) or $HNMe_2$, THF, rt, 1 h (4, 11); o: 8, MeOH, 2–5 h, then $NaBH_4$, 0 °C, 15–30 min; p: NaOMe, MeOH, rt; q: HF · pyridine, EtOAc, rt, 18 h; r: TFA, DCM, 0 °C; s: TMSOTf, THF, 0 °C; t: HF · pyridine, MeCN, rt, 1 week; **d** Depiction of the $^1J_{C-H}$-coupling at the anomeric centre of α-allyl-manQ 4, and of the $^3J_{C-H}$-HMBC-coupling between its anomeric mannose carbon C1 and the C5-hydrogen atom of its cyclopentene moiety. Both couplings were used to confirm the structure of this compound via NMR, and similar couplings were used to confirm the structures of 3, 10 and 11.

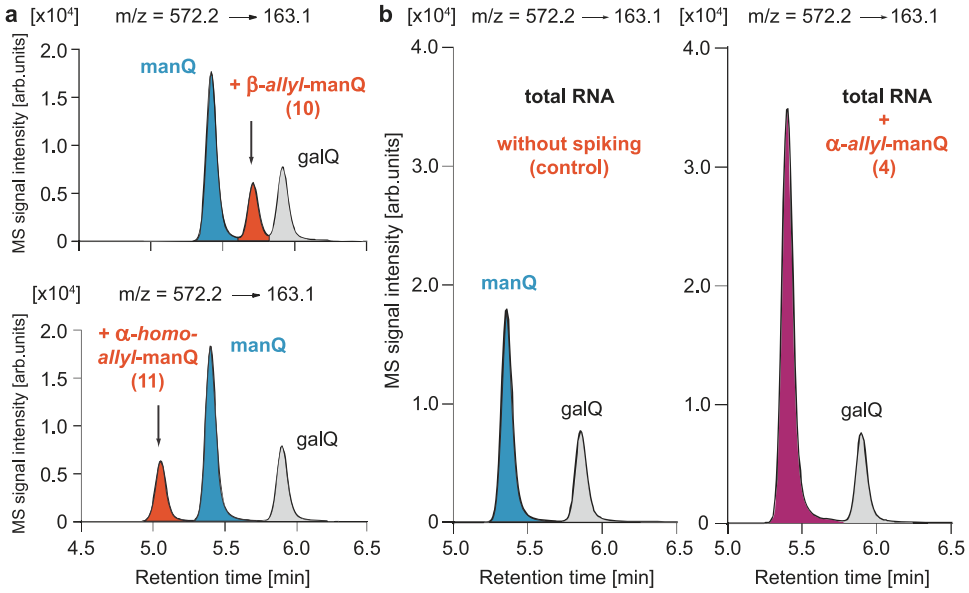


Fig. 5 Chromatograms resulting from the co-injection of the synthetic manQ compounds 10, 11 and 4 with enzymatically digested total RNA from mouse liver analysed by UHPLC-MS/MS. a Co-injection of β-allyl-manQ 10 and α-homoallyl-manQ 11 show the appearance of a third peak, these compounds are therefore different from natural manQ; **b** Co-injection of α-allyl-manQ 4 with digested total RNA from mouse liver shows a complete signal overlap of 4 and natural manQ, thereby leading to an increased signal intensity of manQ (right) in comparison to the control sample without synthetic standard (left). This result shows that our synthetic compound 4 is identical to the naturally occurring manQ nucleoside and establishes its α-allyl-connectivity.

manQ 3. If this reaction was, however, performed at −10 °C for a longer period of time (2 h), we observed formation of the thermodynamically more stable 1-Fmoc-2-TBS-protected compound 15.

With the differently protected cyclopentenes 13, 14 and 15 in hand, the manQ compounds 10, 11 and 4 were synthesised via the glycosylation and reductive amination approach (Fig. 4c). The first target compound, β-allyl-manQ 10, was prepared from cyclopentene precursor 15 starting with a stereoselective β-mannosylation reaction (AgOTf, PhSCl, 2,6-di-tert-butyl-4-methylpyridine, DCM, −78 °C, 2 h, 52%) to give the glycosylated product 17. Fmoc-deprotection (DBU, MeCN, rt, 1 h) and subsequent reductive amination with deazaguanosine precursor 8 gave the fully protected β-allyl-manQ 20. A three-step final deprotection procedure (1: 10% TFA in DCM; 2: HF pyridine, EtOAc, rt, 18 h; 3: NaOMe, MeOH, 2 d) gave β-allyl-manQ 10. For the second target compound, α-homoallyl-manQ 11, we started from cyclopentene precursor 13 and employed Schmidt-Sinaÿ-glycosylation conditions with the literature-known glycosyl donor 16 (TMSOTf, THF, 0 °C, 55%) to obtain glycosylated product 18[36,37]. The synthesis of α-homoallyl-manQ 11 was completed from there by Fmoc-deprotection, reductive amination with 8, and a two-step final deprotection (1: 10% TFA, DCM, 30 min; 2: NaOMe, MeOH, rt, 2 d). The third target compound, α-allyl-manQ 4, was synthesised again via Schmidt-Sinaÿ-glycosylation conditions with donor 16[36,37]. In this case, the SEM-protected cyclopentene precursor 14 was used, because the TBS-protected precursor 15 proved to be sterically too demanding for the glycosylation reaction. Reaction of 14 with 16 however was not fully stereoselective. But after full completion of the synthesis by Fmoc-deprotection, reductive amination and subsequent three-step deprotection (1: NaOMe, MeOH, rt, 18 h; 2: 10% TFA, DCM, 0 °C, 15 min; 3: HF · pyridine, MeCN, rt, 1 week), we obtained a mixture of 10 and 4, which was separated by reversed-phase HPLC-chromatography (see Supplementary Information 3.5). This allowed us to obtain α-allyl-manQ 4 in excellent purity.

We next confirmed the structure of the manQ compounds 4, 10 and 11 using NMR spectroscopy (Fig. 4d). For β-allyl-manQ 10, we observed a $^{1}J_{C\text{-}H}$-coupling at the mannose anomeric centre of 160.2 Hz, indicative of the β-configuration. The allylic connectivity of 10 was confirmed by the $^{3}J_{C\text{-}H}$-HMBC-coupling between the anomeric mannose C1 carbon and the allylic hydrogen C5H of the cyclopentene ring. For α-homoallyl-manQ 11, a $^{1}J_{C\text{-}H}$-coupling of 170.3 Hz was observed at the anomeric centre, proving its α-configuration. In addition, with $\delta = 5.04$ ppm, the chemical shift of the anomeric proton C1H of 11 was as expected higher than the shift observed for the corresponding β-anomer 3 ($\delta = 4.72$ ppm). Homoallylic connectivity of 11 was confirmed by observation of an $^{3}J_{C\text{-}H}$-HMBC-coupling of the anomeric carbon C1 and the homoallylic hydrogen C4H of the cyclopentene moiety. The structure of α-allyl-manQ 4 was proven based on an anomeric $^{1}J_{C\text{-}H}$ -coupling of 170.0 Hz, indicating its α-configuration. Here, the chemical shift of the anomeric proton C1H ($\delta = 4.97$ ppm) was again higher than that of the corresponding β-anomer 10 ($\delta = 4.70$ ppm), thereby also confirming the α-configuration of 4. The allylic connectivity of 4 was confirmed by an $^{3}J_{C\text{-}H}$-HMBC signal resulting from a coupling of the anomeric mannose carbon C1 with the allylic hydrogen C5H of the cyclopentene ring.

With all four possible manQ isomers now available (α/β-allyl-manQ 4 and 10, and α/β-homoallyl-manQ 3 and 11), we next performed individual co-injection studies using again a digest of total RNA from mouse liver as reference material, and the UHPLC-MS/MS-method described above for analysis (Fig. 5). While for β-allyl-manQ 10 and α-homoallyl-manQ 11 LC-MS signals were obtained that were well-separated both from galQ 2 and natural manQ (Fig. 5a), we discovered to our delight a full signal overlap of our synthetic α-allyl-manQ 4 with the natural manQ compound (Fig. 5b) We finally confirmed the overlap of the two compound signals with a second HPLC-MS-method using a different HPLC separation column and gradient (see Supplementary Fig. 3). Here, too, the synthetic material co-eluted with the natural manQ compound. These co-injection experiments

therefore show that the natural manQ compound present in the anticodon loop of $tRNA^{Asp}$ has the chemical structure 4 (see Figs. 1d and 4a). The configuration of its anomeric centre is in fact α, not β. In addition, the mannosyl moiety of natural manQ is connected to the allylic hydroxyl group of the cyclopentene unit, and not to the homoallylic hydroxyl group as in galQ.

To prove that our RNA isolation and digestion conditions do not cause isomerisation of the manQ compounds 3, 4, 10 and 11, we performed stability tests under these conditions with all four isomers (see Supplementary Figs. 4 and 5). In these experiments we observed no isomerisation processes. Finally, we exposed compounds 3 and 4 to neutral (pH 7), basic (0.1 M KOH, pH 13) and acidic (0.1 M HCl, pH 1) conditions for 24 h at 37 °C (see Supplementary Figs. 6 and 7). Even under these enforced settings we did not observe any pH-dependent isomerization processes, indicating that the manQ isomers 3 and 4 are fairly stable compounds.

Based on these experiments, we showed that the widely accepted β-homoallyl structure reported in literature for the non-canonical RNA nucleoside manQ is not correct. Instead, natural manQ is an α-allyl-mannoside (4, Fig. 1d). Its connectivity is therefore maximally different from the β-homoallyl-compound galQ (2, Fig. 1b). The elucidation of the correct manQ structure and the availability of synthetic material will be the basis for illuminating the biological functions of this hypermodified RNA nucleoside.

Methods

Isolation of total RNA from mouse livers. The murine livers for these experiments were kindly provided by Prof. Dr. Stylianos Michalakis. The frozen (−80 °C) livers were pooled and crushed into smaller pieces. The still-frozen pieces were divided into several samples for isolation of their total RNA. To this reason, 1 mL of TriReagent (Sigma Aldrich) was added per 50 mg of still-frozen liver tissue. Then, samples were homogenised with a tissue lyser (Schwingmühle MM400 from Retsch), first at 20 Hz for 4 min, then at 30 Hz for 2 min. Next, 200 µL chloroform per 50 mg of liver tissue were added, and the mixture was centrifugated for full phase separation (12,000 × *g*; 15 min; 4 °C). Of each sample, the aqueous (clear) upper phase was transferred to a new tube and mixed with 500 µL isopropanol per 50 mg of liver tissue. After overnight incubation at −20 °C, samples were centrifugated (21,130 × *g*; 30 min; 4 °C), and the supernatant was carefully removed from the pellets. 1 mL of ice-cold 75% ethanol was added to the pellets, and the samples were vortexed shortly. Again, samples were centrifugated (21,130 × *g*; 20 min; 4 °C), and the supernatant was carefully removed from the pellets. This ethanol washing step was repeated two more times. The resulting total RNA-pellets were first dried at room temperature and then dissolved in nuclease-free water for subsequent enzymatic digestion.

Enzymatic digestion of total RNA to the nucleoside level. 3 µg of mouse (or HEK 293T) total RNA were digested to the nucleoside level using the Nucleoside Digestion Mix (New England BioLabs). To this reason, a solution of 3 µg total RNA in 42.5 µL of nuclease-free water was prepared. 5 µL of the Nucleoside Digestion Mix Reaction Buffer (10x), and 2.5 µL of the Nucleoside Digestion Mix were added, and the mixture was incubated for 2 h at 37 °C. Samples were subsequently supplemented (spiked) with an appropriate amount (see below) of a synthetic manQ compound dissolved in nuclease-free water, or an equal volume of nuclease-free water (control samples). Finally, nuclease-free water was added to give a total sample volume of 100 µL, which therefore was independent of a particular compound concentration. Of note, samples can be stored after digestion and/or spiking for several days at −80 °C without altering the results of subsequent LC/MS analyses.

UHPLC-MS/MS-based co-injection experiments. Co-injection experiments were performed by spiking equimolar (with respect to the natural manQ nucleoside present in the digested total RNA) amounts of heavy-atom-labelled standard to the sample. UHPLC-MS/MS analyses of digested RNA samples were performed using an Agilent 1290 UHPLC system equipped with an UV detector and an Agilent 6490 triple quadrupole mass spectrometer. All samples were filtrated directly before measurement using an AcroPrep Advance 96 filter plate 0.2 µm Supor from Pall Life Sciences. The UHPLC autosampler was cooled to 4 °C. The source-dependent parameters of the MS instrument were as follows: gas temperature 230 °C; gas flow 12 L/min (N2); nebuliser 40 psi; sheath gas heater 300 °C; sheath gas flow 6 L/min (N2); capillary voltage 2.500 V in the positive ion mode; capillary voltage −2.250 V in the negative ion mode; nozzle voltage 0 V. The fragmentor voltage was 380 V. Fragmentation was performed with a collision energy of 35 eV and a cell accelerator voltage of 5 V. A specific fragmentation pattern of m/z = 572.2 → 163.1, as depicted in Supplementary Fig. 1, was observed. Delta EMV was set to 500 (positive mode) and 800 (negative mode). Chromatography was performed using a Poroshell 120 SB-C8 column (Agilent, 2.7 µm, 2.1 mm × 150 mm) at 35 °C with a gradient of water and MeCN, each containing 0.0085% (v/v) formic acid, at a flow rate of 0.35 mL/min. The gradient was as follows: 0 → 8 min, 0 → 2% MeCN (v/v); 8 → 10.9 min, 2 → 3.85% MeCN; 10.9 → 11.3 min, 3.85 → 80% MeCN; 11.3 → 12.0 min, 80% MeCN; 12.0 → 12.3 min, 80 → 0% MeCN; and 12.3 → 14.0 min, 0% MeCN. No other peaks besides the expected ones of manQ, galQ and the spiked standard were detected when using the parameters given above. The identity of galQ was identified by spiking with synthetic galQ-standard (see Supplementary Fig. 2).

Cell culture metabolic feeding experiments. To determine which hexose is incorporated into the natural xQ/manQ base as its sugar moiety, HEK 293T cells were cultured in different heavy RPMI media (see Supplementary Table 1). Each heavy RPMI medium was supplemented with two monosaccharides: One of the two monosaccharides was fully ^{13}C-isotope-labelled, while the other hexose was non-labelled and used in these experiments as a suppressor of metabolic interconversion. Heavy RPMI media with the following hexose combinations were used:

- $^{13}C_6$-D-mannose, D-glucose
- D-mannose, $^{13}C_6$-D-glucose
- $^{13}C_6$-D-galactose, D-Glucose
- D-galactose, $^{13}C_6$-D-glucose

1.3 million HEK 293T cells were seeded on a p60 cell culture dish. Cultivation was performed for 18–21 h in RPMI medium at 37 °C and 5% CO_2. Subsequently, the medium was removed and replaced with heavy RPMI medium. Incubation continued under the same conditions for 30, 60 and 180 min.

Isolation of total RNA from HEK 293T cells. After incubation, the culture medium was removed. The cells were carefully detached from the cell culture dish using PBS and transferred to a 2 mL reaction tube. This was followed by pelleting for 1 min at 500 × *g* and 4 °C. The PBS was removed and the cell pellet immediately resuspended in 1 ml TriReagent (Sigma Aldrich). Samples taken after 30 min in labelled medium were incubated for 40 min, the 60 min- and 180 min-samples were incubated for 5 min. Following incubation, 200 µL of chloroform were added to the sample, which then was heavily vortexed and subsequently centrifuged for 15 min at 12,000 × *g* and 4 °C. The upper clear phase was transferred to a new 2 mL reaction tube. 500 µL of isopropanol were added to the transferred phase and mixed. Precipitation of the RNA was performed at −20 °C overnight. Following this, samples were directly pelleted for 30 min at 21,130 × *g* and 4 °C. The supernatant was carefully removed. Thereafter, 1 ml of 75% (v/v) cold ethanol (−20 °C) was added and the sample was centrifuged for 20 min at 21,130 × *g* and 4 °C. The ethanol washing step including the centrifugation was repeated two more times. The supernatant was removed and the pellet was dried at room temperature. Thereafter, the pellet was dissolved in nuclease-free water.

LC-MS-analysis of total RNA from metabolically labelled HEK cells. HPLC-HESI-MS analysis of the enzymatically digested total RNA of HEK cells (see above) was performed on a Dionex Ultimate 3000 HPLC system coupled to a Thermo Fisher LTQ Orbitrap XL mass spectrometer. Samples were filtrated before the measurement using an AcroPrep Advance 96 filter plate 0.2 µm Supor from Pall Life Sciences. Nucleosides were separated with an Interchim Uptisphere120-3HDO C18 column whose temperature was maintained at 30 °C. Elution buffers were buffer X (2 mM NH_4HCOO in H_2O; pH 5.5) and buffer Y (2 mM NH_4HCOO in H_2O/MeCN 20/80 v/v; pH 5.5) with a flow rate of 0.15 mL/min. The gradient was as follows: 0 → 3.5 min, 0% Y; 3.5 → 4 min, 0 → 0.2% Y; 4 → 10 min, 0.2% Y; 10 → 50 min, 0.2 → 4.7% Y; 50 → 55 min, 4.7 → 60% Y; 55 → 57 min, 60 → 100% Y; 57 → 62 min, 100% Y. The chromatogram was recorded at 260 nm with a Dionex Ultimate 3000 Diode Array Detector, and the chromatographic eluent was directly injected into the ion source of the mass spectrometer without prior splitting. Ions were scanned in the positive polarity mode over a full-scan range of m/z = 210–800 with a resolution of 60,000. Parameters of the mass spectrometer were tuned with a freshly mixed solution of inosine (5 µM) in buffer X and set as follows: Capillary temperature 275 °C; APCI vaporizer temperature 100 °C; sheath gas flow 5.00; aux gas flow 21.0; sweep gas flow 1.00; source voltage 4.80 kV; capillary voltage 0 V; tube lens voltage 45.0 V; skimmer offset 0 V. The ion chromatograms of the compounds of interest were extracted from the total ion current (TIC) chromatogram with a mass range set to ±0.0050 u around the exact mass $[M + H]^+$ of a compound. The peak areas in the extracted ion chromatograms of the heavy and corresponding light compound were integrated and the percentual $^{13}C_6$-labelling of a compound within a sample was calculated.

Synthetic procedures. Synthetic procedures and analytical characterisation of the compounds used in this study are provided in the supporting information.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The HPLC/MS and NMR data generated in this study are provided in the Supplementary Information. Raw data files are available from the corresponding author upon request.

Code availability

For data collection with the TripleQuad LC-MS/MS system, MassHunter Workstation Software Version B.07.01/Build 7.1.524.0 was used. For data collection with the Orbitrap LC-MS-system, the programmes Xcalibur Version 2.1.0 SP1.1160, Chromeleon Version 6.80 SR13 Build 3818 (226197) and LTQ Tune Plus Version 2.5.5 SP2 were used. Analysis of raw data from the TripleQuad LC-MS/MS system was performed with the programme MassHunter Workstation Software Version B.07.01/Build 7.1.524.0. Analysis of raw data from the Orbitrap LC-MS-system was performed with Xcalibur Qual Browser Version 2.1.0 SP1.1160. For general data analysis, Origin 2018 64 bit version was used.

Received: 16 June 2021; Accepted: 11 November 2021;
Published online: 08 December 2021

References

1. Crick, F. Central dogma of molecular biology. *Nature* **227**, 561–563 (1970).
2. Cobb, M. 60 years ago, Francis Crick changed the logic of biology. *PLoS Biol.* **15**, e2003243 (2017).
3. Boccaletto, P. et al. MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res.* **46**, D303–D307 (2018).
4. Cantara, W. A. et al. The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Res.* **39**, D195–D201 (2011).
5. Sarin, L. P. & Leidel, S. A. Modify or die?–RNA modification defects in metazoans. *RNA Biol.* **11**, 1555–1567 (2014).
6. Rozov, A. et al. Novel base-pairing interactions at the tRNA wobble position crucial for accurate reading of the genetic code. *Nat. Commun.* **7**, 10457 (2016).
7. Agris, P. Wobble position modified nucleosides evolved to select transfer RNA codon recognition: a modified-wobble hypothesis. *Biochimie* **73**, 1345–1349 (1991).
8. Yarian, C. et al. Accurate translation of the genetic code depends on tRNA modified nucleosides. *J. Biol. Chem.* **277**, 16391–16395 (2002).
9. Duechler, M., Leszczyńska, G., Sochacka, E. & Nawrot, B. Nucleoside modifications in the regulation of gene expression: focus on tRNA. *Cell. Mol. Life Sci.* **73**, 3075–3095 (2016).
10. Thumbs, P. et al. Synthesis of galactosyl-queuosine and distribution of hypermodified Q-nucleosides in mouse tissues. *Angew. Chem. Int. Ed.* **59**, 12352–12356 (2020).
11. Klepper, F., Jahn, E. M., Hickmann, V. & Carell, T. Synthesis of the transfer-RNA nucleoside queuosine by using a chiral allyl azide intermediate. *Angew. Chem. Int. Ed.* **46**, 2325–2327 (2007).
12. Nishimura, S. Structure, biosynthesis, and function of queuosine in transfer RNA. *Prog. Nucleic Acid Res. Mol. Biol.* **28**, 49–73 (1983).
13. Harada, F. & Nishimura, S. Possible anticodon sequences of tRNAHis, tRNAAsn, and tRNAAsp from Escherichia coli. Universal presence of nucleoside O in the first position of the anticodons of these transfer ribonucleic acid. *Biochemistry* **11**, 301–308 (1972).
14. Katze, J. R., Basile, B. & McCloskey, J. A. Queuine, a modified base incorporated posttranscriptionally into eukaryotic transfer RNA: wide distribution in nature. *Science* **216**, 55–56 (1982).
15. Salinas-Giegé, T., Giegé, R. & Giegé, P. tRNA biology in mitochondria. *Int. J. Mol. Sci.* **16**, 4518–4559 (2015).
16. Kasai, H. et al. Structure of the modified nucleoside Q isolated from Escherichia coli transfer ribonucleic acid. 7-(4,5-cis-Dihydroxy-1-cyclopenten-3-ylaminomethyl)-7-deazaguanosine. *Biochemistry* **14**, 4198–4208 (1975).
17. Morris, R. C., Brown, K. G. & Elliott, M. S. The effect of queuosine on tRNA structure and function. *J. Biomol. Struct. Dyn.* **16**, 757–774 (1999).
18. Meier, F., Suter, B., Grosjean, H., Keith, G. & Kubli, E. Queuosine modification of the wobble base in tRNAHis influences 'in vivo' decoding properties. *EMBO J.* **4**, 823–827 (1985).
19. Tuorto, F. et al. Queuosine-modified tRNAs confer nutritional control of protein translation. *EMBO J.* **37**, e99777 (2018).
20. Zaborske, J. M. et al. A nutrient-driven tRNA modification alters translational fidelity and genome-wide protein coding across an animal genus. *PLoS Biol.* **12**, e1002015 (2014).
21. Marks, T. & Farkas, W. R. Effects of a diet deficient in tyrosine and queuine on germfree mice. *Biochem. Biophys. Res. Commun.* **230**, 233–237 (1997).
22. Rakovich, T. et al. Queuosine deficiency in eukaryotes compromises tyrosine production through increased tetrahydrobiopterin oxidation. *J. Biol. Chem.* **286**, 19354–19363 (2011).
23. Farkas, W. R. Effect of diet on the queuosine family of tRNAs of germ-free mice. *J. Biol. Chem.* **255**, 6832–6835 (1980).
24. Fergus, C., Barnes, D., Alqasem, M. & Kelly, V. The queuine micronutrient: charting a course from microbe to man. *Nutrients* **7**, 2897–2929 (2015).
25. Reyniers, J. P., Pleasants, J. R., Wostmann, B. S., Katze, J. R. & Farkas, W. R. Administration of exogenous queuine is essential for the biosynthesis of the queuosine-containing transfer RNAs in the mouse. *J. Biol. Chem.* **256**, 11591–11594 (1981).
26. Tuorto, F. & Lyko, F. Genome recoding by tRNA modifications. *Open Biol.* **6**, 160287 (2016).
27. Barbieri, I. & Kouzarides, T. Role of RNA modifications in cancer. *Nat. Rev. Cancer* **20**, 303–322 (2020).
28. Suzuki, T. The expanding world of tRNA modifications and their disease relevance. *Nat. Rev. Mol. Cell Biol.* **22**, 375–392 (2021).
29. Kasai, H. et al. Letter: The structure of Q* nucleoside isolated from rabbit liver transfer ribonucleic acid. *J. Am. Chem. Soc.* **98**, 5044–5046 (1976).
30. Okada, N., Shindo-Okada, N. & Nishimura, S. Isolation of mammalian tRNAAsp and tRNATyr by lectin-Sepharose affinity column chromatography. *Nucleic Acids Res.* **4**, 415–423 (1977).
31. Crich, D. & Sun, S. Direct formation of β-mannopyranosides and other hindered glycosides from thioglycosides. *J. Am. Chem. Soc.* **120**, 435–436 (1998).
32. Crich, D., de la Mora, M. & C, R. Synthesis of the mannosyl erythritol lipid MEL A; confirmation of the configuration of the meso-erythritol moiety. *Tetrahedron* **58**, 35–44 (2002).
33. Tvaroska, I. & Taravel, F. R. Carbon-proton coupling constants in the conformational analysis of sugar molecules. *Adv. Carbohydr. Chem. Biochem.* **51**, 15–61 (1995).
34. Okada, N. & Nishimura, S. Enzymatic synthesis of Q nucleoside containing mannose in the anticodon of tRNA: isolation of a novel mannosyltransferase from a cell-free extract of rat liver. *Nucleic Acids Res.* **4**, 2931–2938 (1977).
35. Sharma, V., Ichikawa, M. & Freeze, H. H. Mannose metabolism: more than meets the eye. *Biochem. Biophys. Res. Commun.* **453**, 220–228 (2014).
36. Schmidt, R. R. & Michel, J. Facile synthesis of α- and β-O-glycosyl imidates; preparation of glycosides and disaccharides. *Angew. Chem. Int. Ed.* **19**, 731–732 (1980).
37. Pougny, J.-R. & Sinaÿ, P. Reaction d'imidates de glucopyranosyle avec l'acetonitrile. Applications synthetiques. *Tetrahedron Lett.* **17**, 4073–4076 (1976).

Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement EpiR, No 741912 (T.C.). Further financial support was obtained from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) via SFB1309 (Project-ID 325871075) and SPP 1784 (Project-ID 277203618) (T.C.). We thank the Volkswagen Foundation for generous support through the EvoRib VW-project (T.C.). We thank Prof. Dr. Stylianos Michalakis for kindly providing us with mouse liver samples.

Author contributions

M.H. and P.T. synthesised the compounds. E.K. and M.W. developed UHPLC-MS/MS methods for detection of the manQ derivatives. M.W., E.K. and M.H. performed the LC-MS co-injection-experiments. T.E. performed the metabolic feeding experiments and did the isolation of total RNA from mouse liver and HEK cells. M.W. developed HPLC-MS methods and performed the LC-MS analyses of the feeding experiments. T.C. and M.M. supervised the experiments and the project. T.C. designed the project. T.C. wrote the paper with contributions from M.H., M.W., T.E., E.K. and M.M.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-27371-9.

Correspondence and requests for materials should be addressed to Thomas Carell.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021

4 UNVERÖFFENTLICHTE ARBEITEN

4.1 Untersuchug der intragenomischen Deformylierung mithilfe eines isotopenmarkierten Derivats von 5-Formyldesoxycytidin

Anhand von 2'-fluorierten und isotopenmarkierten Nukleosidderivaten konnte erfolgreich gezeigt werden, dass fdC im genomischen Kontext durch eine Deformylierungsreaktion zu dC demodifiziert werden kann. Dies könnte somit einen Teil der aktiven Demethylierung von genomischem mdC darstellen.[135] Es ist bekannt, dass die Deformylierung von fdC in einer Vielzahl von verschiedenen Zellsystemen auftritt, dabei jedoch unterschiedlich stark ausgeprägt ist.[177] Zur weiteren Untersuchung dieses Prozesses war insbesondere von Interesse, zu welchem Anteil die Deformylierung von fdC im Vergleich zur BER zu der aktiven Demethylierung von mdC beiträgt. Zu diesem Zweck wurde ein isotopenmarkiertes fdC Derivat entwickelt und in zellbiologischen Verfütterungsexperimenten eingesetzt. Es ist bekannt, dass die 2'-fluorierten Formyldesoxycytidin Derivate nach dem Einbau ins Genom nicht im Rahmen von BER ausgeschnitten und durch natürliches dC ersetzt werden können.[183-184] Das nach dem Spalten der glykosidischen Bindung an der Ribose gebildete Oxonium-Ion ist im Falle der fluorierten Derivate so stark destabilisiert, dass die fluorierten Derivate im Genom nicht im Rahmen der BER von Tdg prozessiert werden.[183] Dies bietet den Vorteil, dass man in Verfütterungsstudien dieser Moleküle den direkten Demethylierungsweg durch eine Deformylierung/Decarboxylierung isoliert untersuchen kann. Im Falle des in diesem Kapitel verwendeten isotopenmarkierten $^{13}C_5$-$^{15}N_2$-fdC (**1**) Testmoleküls kann nach dem Einbau ins Genom eine komplexere Fragestellung untersucht werden. Die isotopenmarkierten fdC-Derivate unterscheiden sich in deren chemischen Verhalten nicht signifikant von den natürlich vorkommenden dC-Derivaten. Deswegen ist zu erwarten, dass sie analog zum natürlich vorkommenden fdC und cadC von Tdg erkannt und prozessiert werden. Gleichzeitig können sie jedoch durch eine direkte Deformylierung defunktionalisiert werden. Mit der folgenden Studie sollte untersucht werden, zu welchem Anteil das isotopenmarkierte fdC im Genom deformyliert und wie stark die Deformylierung von fdC in unterschiedlichen Zellsystemen ausgeprägt ist und vom Zelltyp abhängig ist. Im folgenden Kapitel wird die Synthese des isotopenmarkierten Testmoleküls $^{13}C_5$-$^{15}N_2$-fdC (**1**) sowie die Ergebnisse der zellbiologischen Verfütterungsdaten von **1** beschrieben.

Die möglichen Metabolismuswege von **1** nach dem Einbau ins Genom sind in **Abbildung 4.1** dargestellt. Die mögliche aktive Demethylierung mittels BER ist mittels der roten Pfeile hervorgehoben. Hierbei wird unter Beteiligung von Tdg die glykosidische Bindung des

Nukleosids gespalten, wobei eine abasische Stelle entsteht. Diese kann im Anschluss weiter prozessiert und durch das kanonische dC ersetzt werden. Die aktive Demethylierung des Testmoleküls **1** mittels direkter Deformylierung zu $^{13}C_5$-$^{15}N_2$-dC (**2**) ist durch blaue Pfeile markiert.

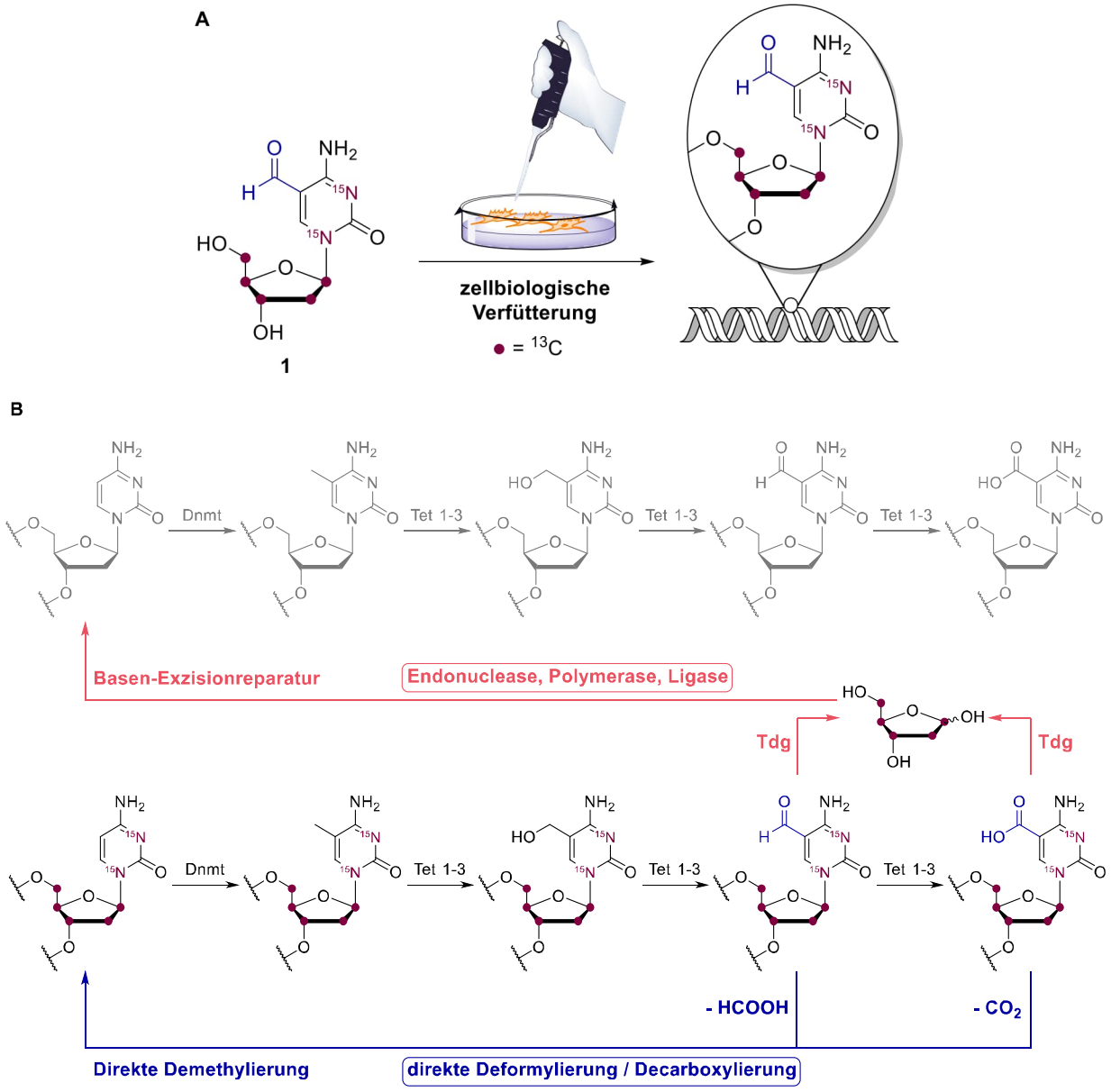
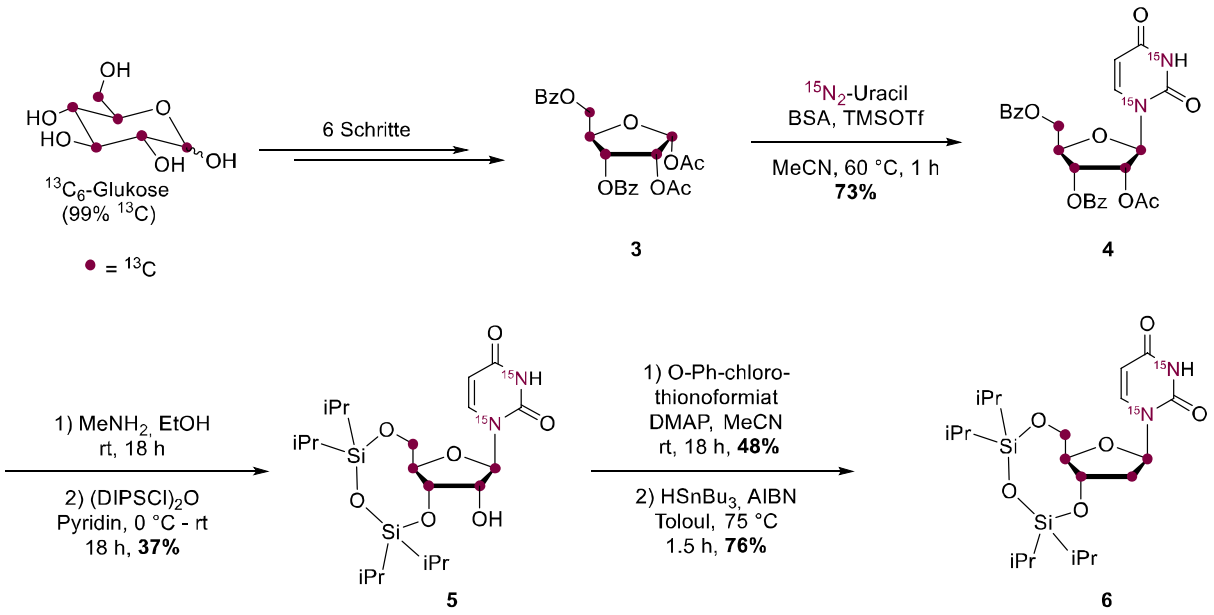


Abbildung 4.1: A: Das Testmolekül **1** wird in zellbiologischen Verfütterungsexperimenten eingesetzt. Nach Einbau ins Genom wird anschließend dessen genomischer Metabolismus untersucht. **B**: Es werden die möglichen Prozessierungswege des verfütterten Testmoleküls **1** dargestellt. Nach dem Einbau ins Genom kann das isotopenmarkierte fdC auf unterschiedlichen Arten weiter prozessiert werden. Einerseits kann es von Tdg erkannt und als Teil der BER nach dem Abspalten der Nukleobase durch ein Desoxycytidin ersetzt werden (dargestellt in rot). Als Teil der aktiven Demethylierung durch eine C-C-Bindungsspaltung kann es direkt zu isotopenmarkierten dC deformylieren und/oder weiter durch Tet-Enzyme zu isotopenmarkiertem cadC oxidiert und im Anschluss decarboxyliert werden (dargestellt in blau).

4.1.1 Synthese von isotopen-markierten Formylcytidin Testmolekülen

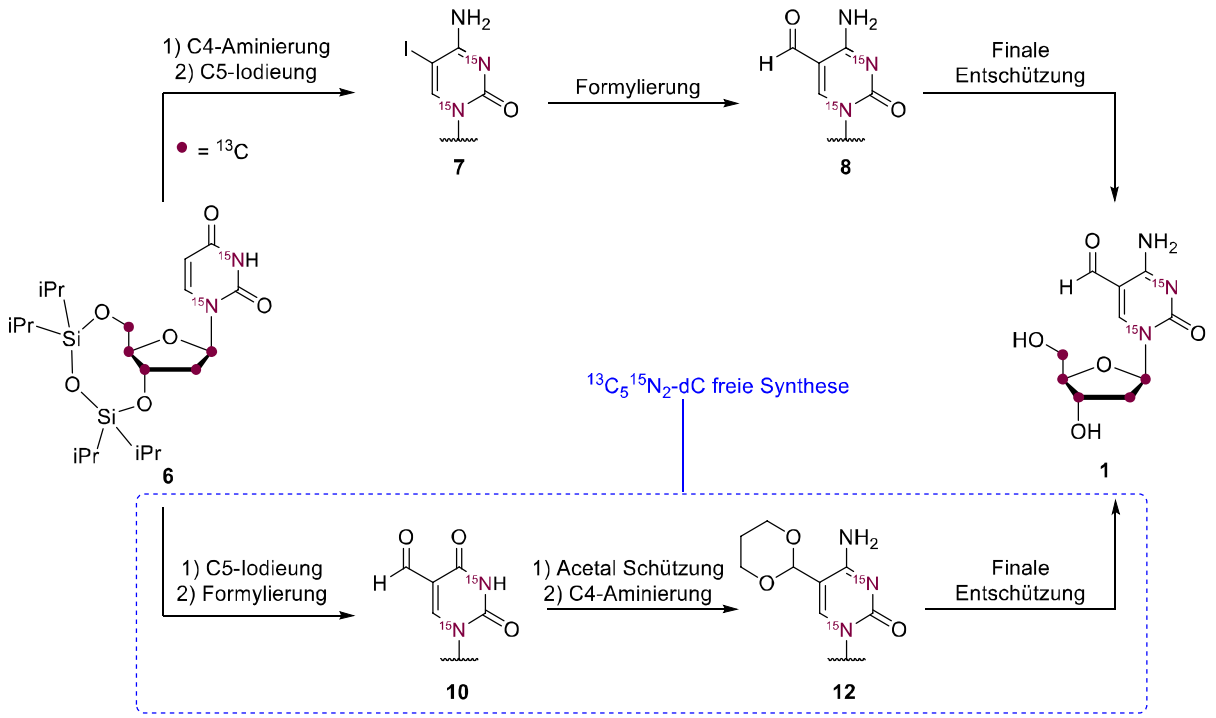
Wie in **Schema 4.1** dargestellt, wurde ausgehend von kommerziell erhältlicher $^{13}C_6$-markierter Glukose in sechs literaturbekannten Schritten das Benzoyl-Acetyl geschützte Ribosederivat **3** hergestellt.[185] Die Acetyl- und Benzoyl-geschützte Ribose **3** stellte ein ideales Vorläufermolekül für die darauffolgende stereoselektive Vorbrüggen-Nukleosidierung mit $^{15}N_2$-Uracil dar. Die nachbargruppeneffekt-gesteuerte Nukleosidierung lieferte ausschließlich das gewünschte siebenfach-isotopenmarkierte β-Uridin **4**. Im nächsten Schritt wurden die Ester-Schutzgruppen unter basischen Bedingungen mithilfe von Methylamin gespalten und die 3',5'-Hydroxygruppen selektiv mit einer Siloxan-Schutzgruppe versehen (**5**). Die 2'-Hydroxygruppe wurde in ein Thionoformiat umgesetzt und in einer Barton-McCombie Desoxygenierung entfernt. So konnte das silyl-geschützte Desoxyuridin **6** erhalten werden.



Schema 4.1: Syntheseschema des siebenfach isotopenmarkierten 3',5'-Siloxan-geschützen Desoxyuridin **6** ausgehend von kommerziell erhältlicher $^{13}C_6$-Glukose.

Wie in Kapitel 3.1 beschrieben war es für die anschließenden zellbiologischen Verfütterungsexperimente zur Untersuchung des komplexen genomischen Prozesses der Deformylierung von fdC essenziell, dass das Probemolekül die höchstmögliche Reinheit aufwies. Insbesondere kleinste Spuren von isotopenmarkiertem dC als Teil einer Verunreinigung im Fütterungsmaterial mussten ausgeschlossen werden, da es ansonsten zum bevorzugten Einbau von dC kommen könnte. Dies würde jegliche Aussagen über eine genomische Deformylierung verfälschen. Aus diesem Grund wurde eine neuartige, $^{13}C_5$-$^{15}N_2$-dC freie Synthese von $^{13}C_5$-$^{15}N_2$-fdC (**1**) entwickelt. Bei der Syntheseroute ist es essenziell, dass $^{13}C_5$-$^{15}N_2$-dC in keinem Syntheseschritt vorkommt oder als Nebenprodukt entstehen

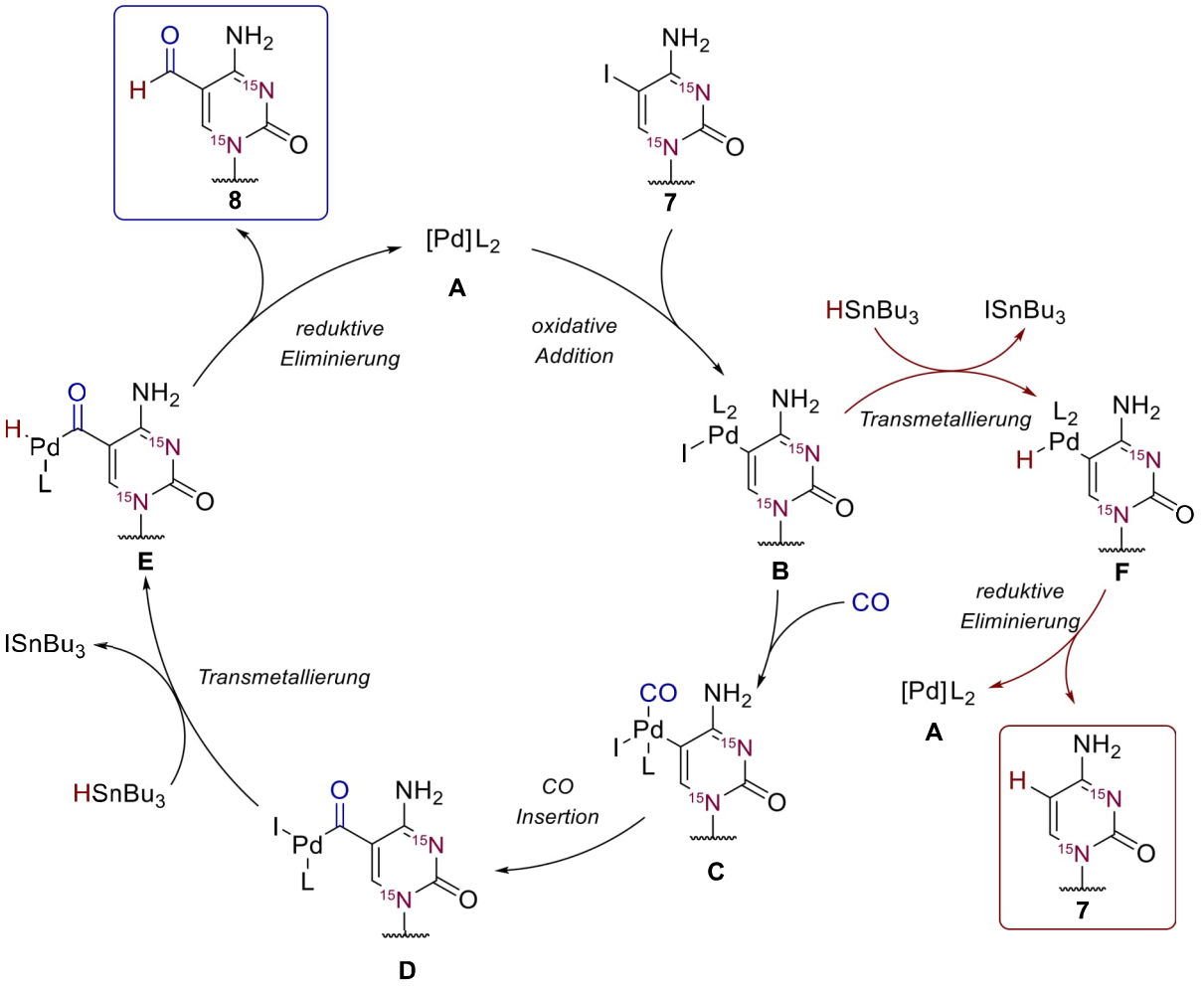
kann. Bei dem im oberen Teil vom **Schema 4.2** dargestellten Syntheseweg, der von *René Rahimoff* entwickelt wurde[135], ist es möglich, dass in der Formylierungsreaktion von **7** zu **8** $^{13}C_5$-$^{15}N_2$-dC als Nebenprodukt entsteht. In dieser Synthesestrategie (**Schema 4.2** oberer Teil) wird das geschützte Desoxyuridin **6** in einer C4-Aminierungsreaktion in das analoge dC-Derivat umgewandelt und im nächsten Schritt an der Base mit Iod zu **8** funktionalisiert. Im Anschluss wird die Schlüsselreaktion, eine Formylierung vom Stille-Typ, durchgeführt, bei der das unerwünschte dehalogenierte Nebenprodukt $^{13}C_5$-$^{15}N_2$-dC (**2**) entstehen kann.



Schema 4.2: Gegenüberstellung zweier verschiedener Synthesestrategien zur Darstellung des Testmoleküls **1** ausgehend von geschütztem Desoxyuridin-Derivat **6**. Der obere Syntheseweg verläuft über das dC-Derivat **7**, das in einer Stille-artigen Formylierung in **8** umgewandelt wird. Im unteren Teil ist die neuartige Synthesestrategie gezeigt, die frei von $^{13}C_5$-$^{15}N_2$-dC verläuft, um mögliche Kontaminationen auszuschließen.

Die Dehalogenierungsreaktion ist eine gängige Nebenreaktion der Stille-artigen Formylierung und ist mechanistisch in **Schema 4.3** dargestellt. Im bevorzugten und erwünschten Hauptkatalysezyklus erfolgt eine oxidative Addition von iodiertem Nukleosid **7** an das Pd(0) zu **B**. Anschließend kommt es zur Koordination von CO an **C** und es findet eine Insertion von CO in die Pd-C5-Nukleosid Bindung zu **D** statt. Eine Transmetallierung des Hydrid-Ions von $HSnBu_3$ auf Pd führt zu Spezies **E,** welche im Anschluß eine reduktive Eliminierung durchläuft und somit das formylierte Nukleosid **8** liefert. Der Palladium-Katalysator wird schließlich regeneriert zu **A**. Neben diesem Hauptkatalysezyklus kann ein Transmetallierungsschritt des $HSnBu_3$ direkt ausgehend von der Spezeis **B** stattfinden. Dies tritt vor allem dann auf, wenn die Insertion von CO in die Pd-Nukleosidbindung nicht schnell stattfindet. In diesem

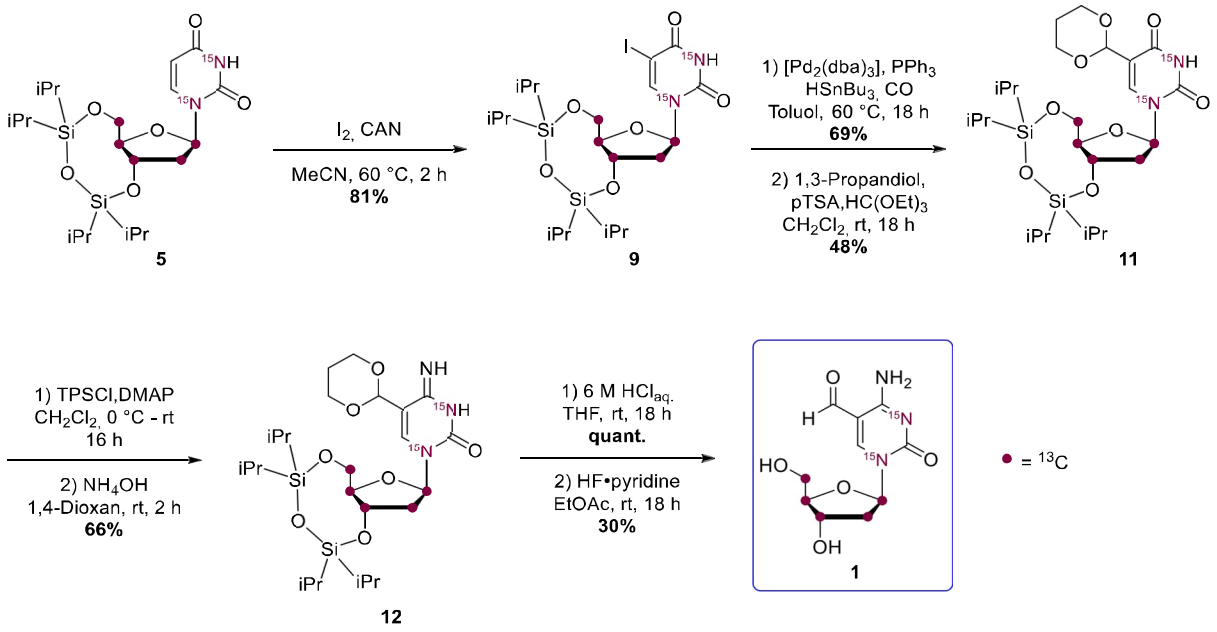
unerwünschten Schritt wird Spezies **F** gebildet, die nach einer reduktiven Eliminierung das dehalogenierte Desoxycytidin-Derivat **2** als Nebenprodukt dieser Reaktion liefert.



Schema 4.3: Katalysezyklus der carbonylativen Stille Reaktion von **7** zu **8**, bei der es zur Bildung von unerwünschtem Nebenprodukt **2** kommen kann. Unter den hier verwendete Bedingungen ist L = PPh_3.

Um das Vorkommen von $^{13}C_5$-$^{15}N_2$-dC (**2**) in der Syntheseroute absolut auszuschließen, habe ich eine alternative $^{13}C_5$-$^{15}N_2$-dC-freie Syntheseroute entwickelt (**Schema 4.4**), in dem alle Zwischenstufen der Synthese aus dem $^{15}N_2$-$^{13}C_5$-Desoxyuridin Analogon verlaufen. Diese Syntheseroute, bei der das $^{13}C_5$-$^{15}N_2$-Desoxyuridin im letzten Schritt zum $^{13}C_5$-$^{15}N_2$-Desoxycytidin Derivat **12** umgewandelt wird, ist im unteren Teil vom **Schema 4.2** dargestellt. Bei der neuen Synthesestrategie wird das silyl-geschützte Uridin **6** direkt an C-5 mit Iod funktionalisiert (**9**) und im Anschluss mit CO formyliert zu **10**. Der entsprechende Aldehyd wird in Form eines Acetals geschützt (**11**) und im nächsten Schritt wird durch eine Aminierung an C-4 das funktionalisierte Desoxyuridin-Derivat **11** in das analoge Desoxycytidin **12** umgewandelt. Durch eine anschließende Abspaltung der Schutzgruppen kann so in einem dC-freien Syntheseweg das isotopenmarkierte $^{13}C_5$-$^{15}N_2$-fdC (**1**) erhalten werden.

In der $^{15}N_2$-$^{13}C_5$-dC freien Synthese wurde das silyl-geschützte Desoxyuridin **6** regioselektiv mit molekularem Iod und Ammonium-Cer(IV) Nitrat an der C-5 Position der Nukleobase funktionalisiert (**9**) und so der ideale Vorläufer für eine Pd-katalysierte C-C-Kupplung dargestellt. Anschließend wurde das Nukleosid unter CO-Atmosphäre in einer Stille-artigen Kreuzkupplung formyliert und im Anschluss der Aldehyd als Acetal geschützt (**11**). Eine Aminierung mit Triisopropylsulfonylchlorid und anschließender Behandlung mit wässrigem Ammoniak führte zum Desoxycytidin Derivat **12**. Saure Abspaltung der Acetal-Schutzgruppe und Spaltung der Silylschutzgruppen mit HF·Pyridin führte nach HPLC-Aufreinigung in guten Ausbeuten und exzellenter Reinheit zum siebenfach isotopenmarkierten fdC Derivat $^{13}C_5$-$^{15}N_2$-fdC (**1**). Der Syntheseweg ist in **Schema 4.4** dargestellt.



Schema 4.4: Syntheseschema des siebenfach isotopenmarkierten Testmoleküls $^{13}C_5$-$^{15}N_2$-fdC (**1**) ausgehend von silylgeschütztem Desoxyuridin **6**.

4.1.2 Zellbiologische Daten des verfütterten Testmoleküls

Nachdem das im vorherigen Kapitel beschriebene Testmolekül $^{13}C_5$-$^{15}N_2$-fdC (**1**) erfolgreich über eine $^{13}C_5$-$^{15}N_2$-dC-freie Synthese dargestellt wurde und eine exzellente Reinheit aufwies, wurde es (von *Angie Kirchner* und *Ewelina Kamińska*) in zellbiologischen Verfütterungsexperimenten verwendet. In Anwesenheit des Testmoleküls **1** wurden Zellkulturen von embryonalen Stammzellen von Mäusen sowie unterschiedliche somatische und Krebszellarten für 72 h kultiviert. Dieser experimentelle Zeitrahmen war ausreichend, damit das Testmolekül ins entsprechende Genom eingebaut werden konnte sowie weitere metabolische Prozesse durchlaufen konnte. Im Anschluss wurden die Zellen geerntet, deren DNA isoliert und mittels einer enzymatischen Verdaumethode in die einzelnen Nukleoside

aufgespalten. Das resultierende Nukleosidgemisch wurde mit isotopenmarkierten internen Standards versetzt und anschließend mittels UHPLC-MS/MS analysiert. Mit diesen Experimenten sollte die Einbaueffizienz von **1** ins Genom untersucht werden, sowie die Frage beantwortet werden, wie stark die genomische Deformylierung von $^{13}C_5$-$^{15}N_2$-fdC (**1**) zu $^{13}C_5$-$^{15}N_2$-dC (**2**) ausgeprägt ist. **Abbildung 4.2** zeigt schematisch einen Teil der verschiedenen untersuchten biologischen Prozesse, die bei diesem Experiment analysiert werden.

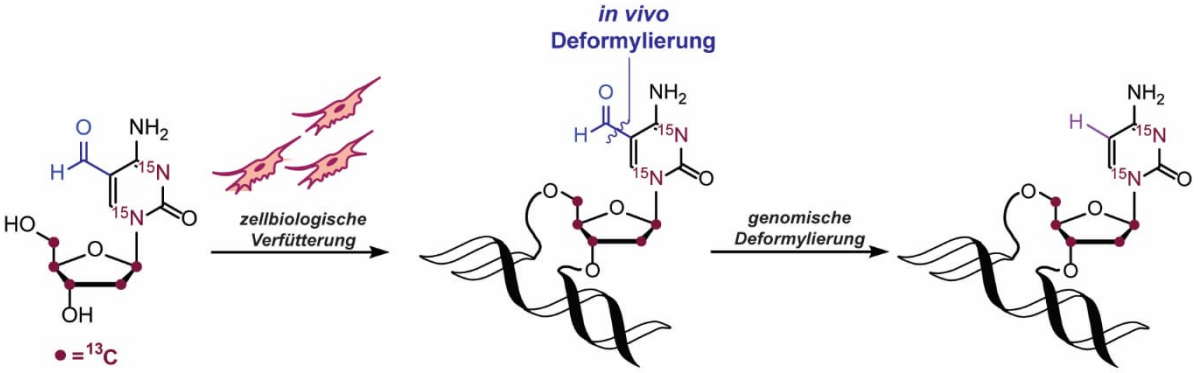


Abbildung 4.2: Darstellung der biologischen Prozesse, welche das siebenfach isotopenmarkierteTestmolekül $^{13}C_5$-$^{15}N_2$-fdC (**1**) im experimentellen Rahmen durchläuft. Nach der zellbiologischen Verfütterung wurde das Testmolekül von den Zellen ins Genom eingebaut. Nach anschließender genomischer Deformylierung wird so isotopenmarkiertes $^{13}C_5$-$^{15}N_2$-dC (**2**) gebildet.

Basierend auf früheren Studien mit den 2'-fluorierten dC Derivaten[135, 177], wurde als erstes zellbiologisches System die mausembryonale Stammzelllinie K3 verwendet. Die aus dem oben beschriebenen experimentellen Setup resultierenden massenspektrometrischen Daten zeigen, wie in **Abbildung 4.3** dargestellt, dass das isotopenmarkierte Testmolekül $^{13}C_5$-$^{15}N_2$-fdC (**1**) in absoluten Mengen von $5.3x10^{-4}$/dN im Genom der Zellen zu finden ist. Dies entspricht einer Einbaueffizienz von etwa 20%, was bedeutet, dass rund 20% des genomischen fdC als markiertes $^{13}C_5$-$^{15}N_2$-fdC (**1**) vorliegt. Interessanterweise konnte man in diesem Fall keine Deformylierung von $^{13}C_5$-$^{15}N_2$-fdC zu $^{13}C_5$-$^{15}N_2$-dC oder eine anschließende Remethylierung von $^{13}C_5$-$^{15}N_2$-dC zu $^{13}C_5$-$^{15}N_2$-mdC auf genomischer Ebene beobachten (**Abbildung 4.3**). Die Level dieser Modifikationen lagen in diesem Experiment unter dem Detektionslimit der Messmethode (< LOD). Diese Beobachtung zeigt, dass das isotopenmarkierte Testmolekül **1** im löslichen Nukleosidpool von Kinasen phosphoryliert wird und das entsprechende Nukleotid von Polymerasen erfolgreich ins Genom eingebaut wird. Eine genomische Deformylierung konnte in diesem Fall jedoch nicht beobachtet werden. Dies könnte darauf zurückzuführen sein, dass die Rate der genomischen Deformylierung von fdC stark von der Art und Herkunft des Zellsystems abhängt, wie in Kapitel 3.1 beschrieben wurde. Möglicherweise ist in diesem Fall keine Deformylierung zu beobachten, da dieser Prozess in der K3 Zelllinie nicht sehr stark ausgeprägt ist und somit die Level an $^{13}C_5$-$^{15}N_2$-dC unterhalb

des Detektionslimits liegen. Weitherhin ist es möglich, dass das ins Genom eingebaute $^{13}C_5$-$^{15}N_2$-fdC entweder stabile Level im Genom erreicht oder an anderen genomischen Prozessen wie BER beteiligt ist.

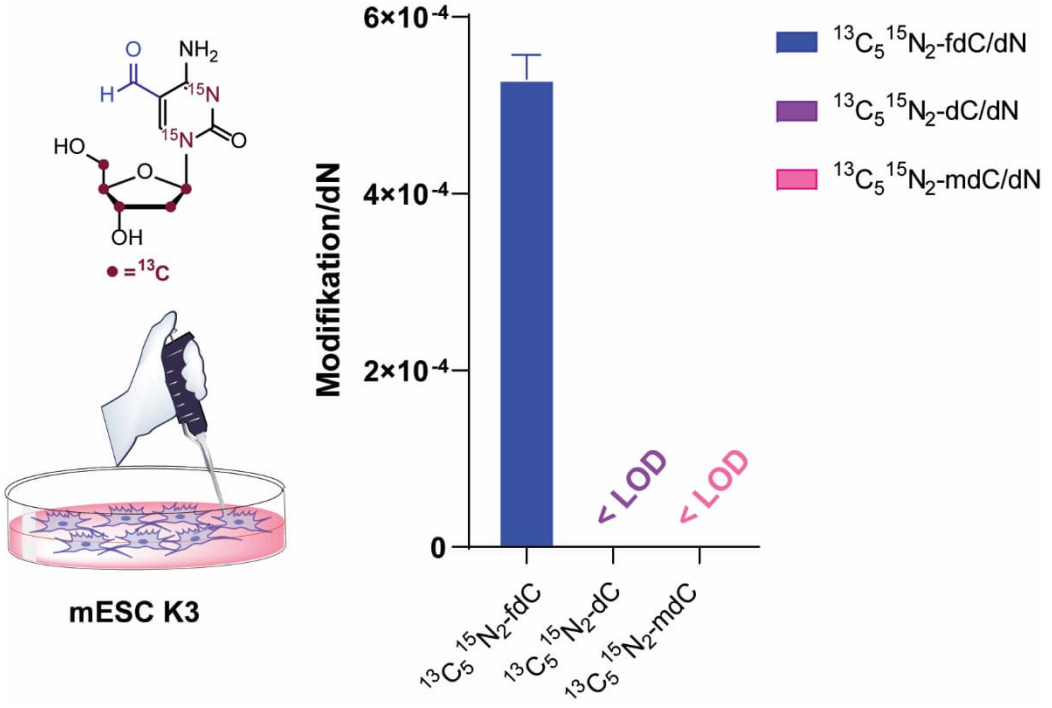


Abbildung 4.3: Darstellung der experimentellen Level der verschiedenen isotopenmarkierten dC-Derivaten im Genom nach zellbiologischer Verfütterung von $^{13}C_5$-$^{15}N_2$-fdC (**1**) an die mESC K3 Zelllinie (LOD = *limit of detection*; Detektionslimit der Messmethode; < LOD = unter dem Detektionslimit).

Im Gegensatz zu 2'-fluoriertem fdC, das im Genom der K3 mESCs zum erheblichen Teil in einer Deformylierungsreaktion zu 2'-F-dC umgesetzt wird[135, 177], kann das isotopenmarkierte fdC, dass dem natürlich vorkommenden fdC chemisch (bezogen auf Reaktivität) identisch ist, weiteren genomischen Prozessen unterliegen. Das eingebaute isotopenmarkierte Testmolekül **1** kann analog zum natürlich vorkommenden genomischen fdC von Tdg erkannt und durch die Basenexzisionsreparatur-Maschinerie aus dem Genom entfernt werden. Wie **Abbildung 4.4** zeigt, kann neben der Deformylierung von $^{13}C_5$-$^{15}N_2$-fdC durch einen direkten C-C-Bindungsbruch ein Bruch der glykosidischen Bindung durch Tdg erfolgen. Dies führt zur Bildung einer abasischen Stelle, welche im Folgenden über eine β-Eliminierung zu einer aufgebrochenen DNA führen kann oder durch ein unmodifiziertes kanonisches dC ersetzt werden kann und somit nicht direkt detektierbar ist. In einer früheren Studie wurde jedoch gezeigt, dass die Anwesenheit von fdC und cadC zu keiner Akkumulation von schädlichen Strangbrüchen führt.[123] Beide Prozesse, sowohl die direkte DNA-Demethylierung als auch die Basenexzisionsreparatur, können gleichzeitig stattfinden. Deren Verhältnisse sind jedoch stark von der untersuchten Zellart abhängig. Deshalb wurden in einer Reihe von Experimenten zum einen unterschiedlichen Zelltypen miteinander verglichen und zum anderen verschiedene

spezielle Zelllinien untersucht, bei denen die Aktivität bestimmter Enzyme, welche einen Einfluss auf diese Prozesse aufweisen, ausgeschaltet wurde (*Knockout* Zelllinien).

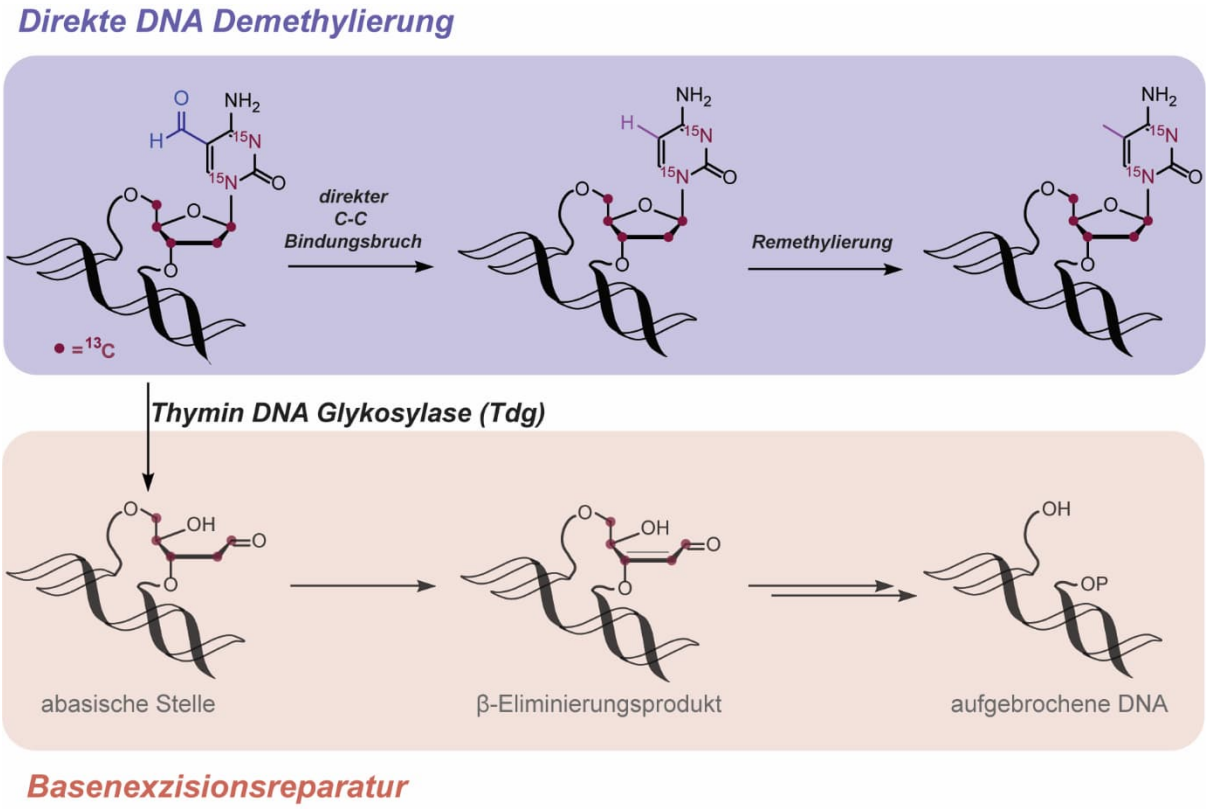


Abbildung 4.4: Zwei mögliche Demethylierungswege von $^{13}C_5$-$^{15}N_2$-fdC. Das ins Genom eingebaute isotopenmarkierte fdC kann über einen direkten C-C-Bindungsbruch defunktionalisiert und zum isotopenmarkierten dC umgewandelt werden (oberer Teil – Direkte DNA-Demethylierung). Gleichzeitig wird das isotopenmarkierte fdC von Tdg erkannt und kann durch eine Spaltung der glykosidischen Bindung zu einer abasischen Stelle, bis hin zu einer aufgebrochenen DNA führen (unterer Teil – Basenexzisionsreparatur).

Um zu untersuchen, in welchem Rahmen die Tdg-Aktivität Einfluss auf den Einbau und die Deformylierung von fdC ausübt, wurde das Verfütterungsexperiment in einer Tdg-*Knockout* Zelllinie durchgeführt. **Abbildung 4.5** stellt die quantitativen Daten nach der Verfütterung von **1** an Wildtyp-Zellen (wt) im Vergleich zu Tdg-KO Zellen dar. Wie erwartet, ist bei der absoluten, ins Genom eingebauten Menge an $^{13}C_5$-$^{15}N_2$-fdC (**1**) eine deutliche Erhöhung von 65% in Tdg-KO-Zellen im Vergleich zum Wildtyp zu beobachten. Interessanterweise ist bei den ermittelten Mengen an $^{13}C_5$-$^{15}N_2$-dC, welches aus der Deformylierungsreaktion von isotopenmarkiertem fdC entsteht keine Veränderung zu sehen. Sowohl im Wildtyp, als auch in Tdg-*Knockout*-Zellen wurde kein $^{13}C_5$-$^{15}N_2$-dC (**2**) detektiert. Dieses Ergebnis deutet darauf hin, dass die nicht detektierbare und damit sehr niedrige Deformylierungsaktivität der K3 mESCs bei der Verfütterung von **1** unabhängig von der Aktivität von Tdg ist und von anderen Faktoren abhängt. Bei der Untersuchung des Genoms nach den Verfütterungsexperimenten wurde ein weiteres isotopenmarkiertes Nukleosid detektiert. Wie aus der **Abbildung 4.5** zu entnehmen

ist, wurde im Genom der K3 Wildtyp- als auch in den Tdg-KO-Zellen $^{13}C_5$-$^{15}N_2$-fdU (**13**) in absoluten Mengen von 1.5 – 1.7 x10^{-5}/dN detektiert. Im Vergleich von Wildtyp mit Tdg-KO-Zellen wiesen die Tdg-KO-Zellen 10% mehr an $^{13}C_5$-$^{15}N_2$-fdU (**13**) im Vergleich zum Wildtyp auf. Es wird vermutet, dass $^{13}C_5$-$^{15}N_2$-fdU (**13**) im löslichen Nukelosidpool der Zellen mithilfe von Desaminasen aus $^{13}C_5$-$^{15}N_2$-fdC (**1**) durch eine Desaminierungsraktion gebildet wird. Diese Desoxyuridin-Modifikation kann anschließend ins Genom eingebaut werden.[186] Es ist bekannt, das Thymidin und Thymidilat Kinasen Uridin-Modifikationen phosphorylieren können, die dann im Anschluss ins Genom eingebaut werden.[187] Die relativ hohen Menge an dieser möglicherweise toxischen dU-Modifikation könnte eine mögliche Erklärung für die nicht messbare Deformylierungsaktivität in diesen Zellsystemen darstellen. Um die genomische Deformylierung von fdC genauer zu untersuchen, wurden für weitere Experimente weitere Zelltypen, die sich in ihrer Zellteilungsrate und Herkunft unterscheiden, getestet.

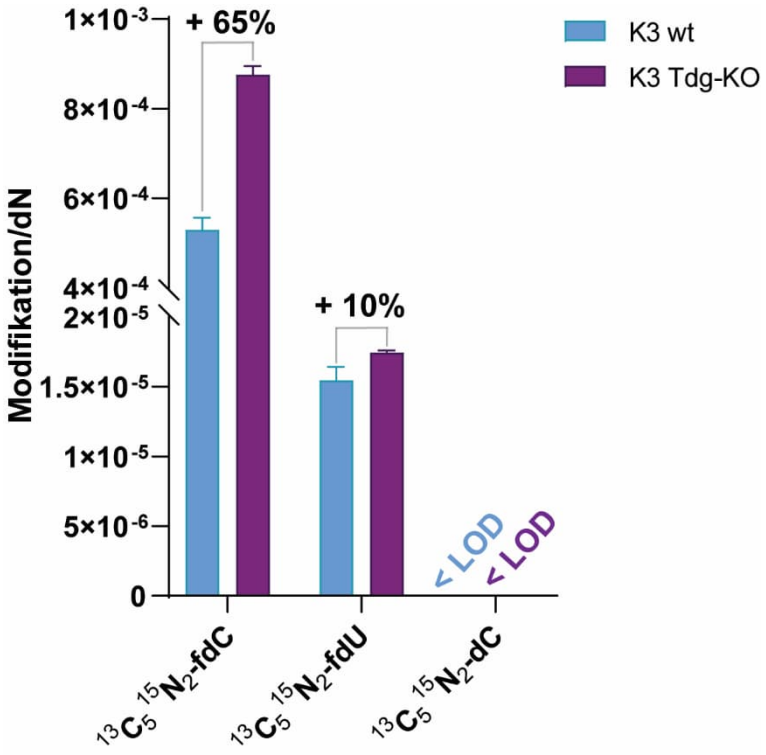


Abbildung 4.5: Absolute Level von unterschiedlichen isotopenmarkierten Nukelosiden detektiert im Genom von K3 Wildtyp- und K3 Tdg-*Knockout* Zellen nach Verfütterung von **1**.

Hierzu wurde $^{13}C_5$-$^{15}N_2$-fdC (**1**) im Zellmedium unterschiedlicher Zelltypen für 72 h inkubiert. Bei den hier verwendeten Zelllinien handelt es sich um mausembryonale Stammzellen, mESC J1, mESC K3 und mESC E14 Zellen, somatische Zellen (CHO-K1), sowie Krebszellen, die aus humanen Darmkrebszellen erhalten wurden (HCT-116). Aus früheren Studien (Kapitel 3.1) ist bekannt, dass es im Falle von 2'-fluorierten Testmolekülen nach deren zellbiologischer Verfütterung zu erheblichen Unterschieden in der Deformylierungsrate in unterschiedlichen

Zellsystemen kommt. Deshalb sollte in diesem Zusammenhang die Deformylierung in möglichst vielen unterschiedlichen Zelltypen untersucht werden. Nach der Verfütterung von **1** wurde die Menge des eingebauten $^{13}C_5$-$^{15}N_2$-fdC (**1**) sowie des Deformylierungsproduktes $^{13}C_5$-$^{15}N_2$-dC (**2**) in der genomischen DNA der Zellen quantifiziert. Wie aus den Daten in **Abbildung 4.6** ersichtlich ist, gibt es abhängig vom Zelltyp große Unterschiede in der Menge von isotopenmarkierten dC-Derivaten, die sich im Genom befinden. Die mESC K3 Zellen wiesen die höchsten Mengen an $^{13}C_5$-$^{15}N_2$-fdC/dN auf, während in der HCT-116 Zelllinie die niedrigsten Werte beobachtet wurden. Interessant ist, dass in den K3 Zellen hauptsächlich $^{13}C_5$-$^{15}N_2$-fdC/dN zu finden war und nur eine vergleisweise geringe Desaminierungsaktivität zu $^{13}C_5$-$^{15}N_2$-fdU detektiert wurde. Im Allgemeinen wurde in dieser Zelllinie eine eher geringe Einbaurate isotopenmarkierter Nukleoside ins Genom beobachtet. Im Kontrast dazu wies die mausembryonale Stammzelllinie J1 eine viel höhere Einbaurate an isotopenmarkierten Nukleosiden auf, wobei jedoch auch eine hohe Desaminierungsaktivität zu beobachten war, was die hohe Einbaurate von $^{13}C_5$-$^{15}N_2$-fdU ins Genom belegt. Gleichzeitig konnte allerdings in der J1-Zelllinie eine eindeutige genomische Deformylierung von $^{13}C_5$-$^{15}N_2$-fdC (**1**) zu $^{13}C_5$-$^{15}N_2$-dC (**2**) gezeigt werden, da quantifizierbare Mengen von $6.5x10^{-5}$/dN an $^{13}C_5$-$^{15}N_2$-dC (**2**) im Genom detektiert wurden. Ein vergleichbares Bild zeigt die Verfütterung von $^{13}C_5$-$^{15}N_2$-fdC (**1**) an die E14 Zellen, bei denen eine ähnliche Einbaueffizienz sowie Deformylierungsrate zu beobachten war. Auch hier wurden relativ hohe Mengen von $1.0x10^{-3}$/dN an $^{13}C_5$-$^{15}N_2$-fdU im Genom detektiert. Interessanterweise zeigten die untersuchten somatischen Zellen, CHO-K1 sowie die Krebszelllinie HCT-116, basierend auf den detektierten Menge von isotopenmarkierten dC-Derivaten ein stark unterschiedliches Verhalten auf. Das verfütterte Testmolekül **1** wurde im Genom detektiert, jedoch konnte keine oder nur eine sehr geringe Deformylierungsaktivität beobachtet werden. Die Level an $^{13}C_5$-$^{15}N_2$-fdU (**13**) im Genom betrugen zwischen 1.3 – $1.4x10^{-4}$/dN und wiesen somit eine sehr viel niedrigere Desaminierungsaktivität im Vergleich zu den hier untersuchten mESCs auf. Aus diesen Daten ist klar ersichtlich, dass sich die Prozessierung von epigenetisch relevantem fdC im Genom durch eine Deformylierungsreaktion je nach Zelltyp aufgrund deren spezifischen Eigenschaften stark unterscheidet. Dies deutet darauf hin, dass die Deformylierungsreaktion ein regulierter Prozess ist, welcher in Abhängigkeit von dem Zustand des untersuchten biologischen Systems stark variiert.

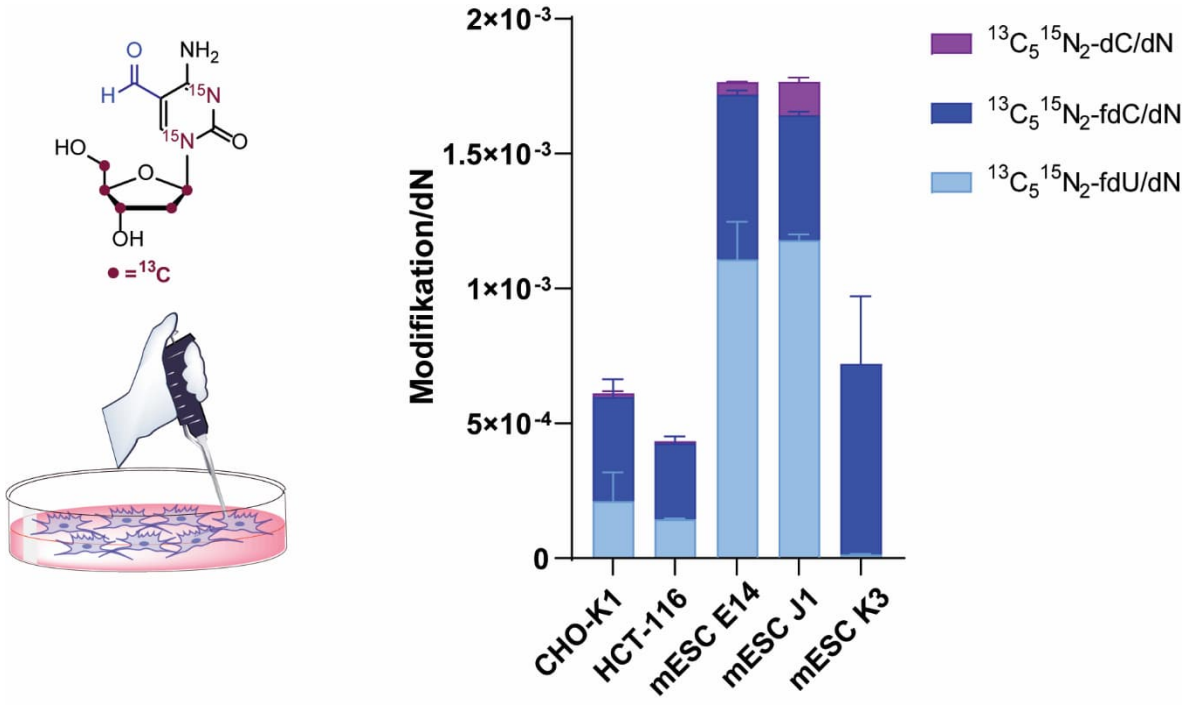


Abbildung 4.6: Absolute Level von unterschiedlichen isotopenmarkierten Nukelosiden detektiert im Genom von unterschiedlichen Zellarten nach Verfütterung von **1**. Es wurden folgende Zelltypen untersucht: CHO-K1, HCT-116, mESC E14, mESC J1, mESC K3.

Zur Ermittlung der idealen Bedingungen zur Untersuchung der Deformylierungsreaktion sowie der maximalen Einbaueffizienz, wurde das Testmolekül **1** in unterschiedlichen Konzentrationen im Bereich von 0.5 µM bis 100 µM in Verfütterungsexperimenten eingesetzt. Hierzu wurde **1** dem Zellmedium der E14 Zellen zugesetzt und anschließend für 72 h inkubiert. Im Folgenden wurde analog zu den vorherigen Experimenten die genomische DNA isoliert und mittels massenspektrometrischer Methoden untersucht. Die in **Abbildung 4.7 A** dargestellte *Labelling* Effizienz beschreibt die Menge an genomischem fdC im Verhältnis zu der Menge des in das Genom eingebauten isotopenmarkierten Testmoleküls **1**. Wie aus der **Abbildung 4.7 A** zu entnehmen ist, wird eine Mindestkonzentration an **1** von 2 µM im Zellmedium benötigt, damit das fdC-Derivat in detektierbaren Mengen ins Genom eingebaut wird. Bei dieser Konzentration liegt die *Labelling* Effizienz bei > 5%. Durch die Erhöhung der Konzentration im Verfütterungsexperiment ist ein linearer Anstieg der Einbaueffizienz zu sehen. Bei der höchsten getesteten Konzentration von 100 µM **1** stellen mehr als 20% aller genomischen fdCs die isotopenmarkierte $^{13}C_5$-$^{15}N_2$-fdC Variante dar. Eine weitere Konzentrationserhöhung auf über 100 µM erwies sich als kontraproduktiv, da **1** in einem höheren Konzentratrionsbereich für die Zellen hochtoxisch ist. Die höchste Einbaueffizienz von > 20% und dadurch die idealen Bedingungen für die Untersuchung der genomischen Deformylierung von $^{13}C_5$-$^{15}N_2$-fdC, wurden also mit einer Konzentration von **1** von 100 µM im zellulären Medium

erreicht. Die absoluten Mengen von **1**, die ins Genom eingebaut wurden, sind in **Abbildung 4.7 B** gezeigt.

Wie in **Abbildung 4.7 A** gezeigt ist zu erkennen, dass die Menge von eingebautem Material proportional zur Konzentration des Testmoleküls im Zellmedium ansteigt. Ab einer Konzentration von 5 μM ist außerdem die Präsenz von $^{13}C_5$-$^{15}N_2$-dC (**2**) im Genom detektierbar, welches durch eine Deformylierungsreaktion von **1** ensteht. Die Menge des Deformylierungsproduktes steigt jedoch nicht proportional zur Menge von **1** im Genom, sondern liegt bei allen untersuchten Konzentrationen (> 5 μM) im ähnlichen Bereich. Die Deformylierungsrate scheint ein *steady state* zu erreichen, das heißt, nur ein bestimmter Teil von $^{13}C_5$-$^{15}N_2$-fdC (**1**) wird im Genom zu $^{13}C_5$-$^{15}N_2$-dC (**2**) deformyliert. Das verbleibende genomische $^{13}C_5$-$^{15}N_2$-fdC (**1**) bleibt möglicherweise im Genom als stabiler epigenetischer Marker vorhanden. Es wird angenommen, dass das eingebaute **1** zufällig in der DNA verteilt ist und nicht nur in den epigenetisch relevanten CpG Inseln vorkommt, da der Einbau zufällig stattfindet. Die Beobachtung, dass unabhängig von der verfütterten **1** Konzentration nur ein Teil von $^{13}C_5$-$^{15}N_2$-fdC (**1**) zu $^{13}C_5$-$^{15}N_2$-dC (**2**) deformyliert wird, könnte dadurch erklärt werden, dass nur das sich in den CpG Inseln befindliche $^{13}C_5$-$^{15}N_2$-fdC (**1**) durch einen Prozess der direkten Demethylierung zu $^{13}C_5$-$^{15}N_2$-dC (**2**) umgesetzt wird. Sollte die Deformylierung ein enzymatischer Prozess sein, der zur Regulierung der Mengen der epigenetisch relevanten Nukleoside dient, wäre naheliegend, dass ein solcher Prozess an genau definierten Stellen im Genom stattfindet. Neben den genomischen Mengen von $^{13}C_5$-$^{15}N_2$-fdC (**1**) und $^{13}C_5$-$^{15}N_2$-dC (**2**) ist in der **Abbildung 4.7 B** auch die Menge des detektierten $^{13}C_5$-$^{15}N_2$-fdU (**13**) zu sehen. Die detektierten Mengen steigen proportional zur Konzentration des isotopenmarkierten Fütterungsmaterials und erreichen Level von bis zu $1x10^{-3}$/dN bei 100 μM **1** im Zellmedium. Der Einbau dieser Modifikation scheint effizienter zu sein als der Einbau der entsprechenden Cytidin-Modifikation. Wie später detailliert beschrieben wird ist dies ein unerwünschter Prozess, da ein Einbau von großen Mengen an $^{13}C_5$-$^{15}N_2$-fdU (**13**) durch die Toxizität dieser Modifikation zum Absterben der beteiligten Zellen führt. Diese Daten zeigen jedoch eindeutig, dass eine genomische Deformylierung von $^{13}C_5$-$^{15}N_2$-fdC (**1**) zu $^{13}C_5$-$^{15}N_2$-dC (**2**) stattfindet. Die Rate dieses Prozesses scheint nicht von der Menge des eingebauten $^{13}C_5$-$^{15}N_2$-fdC (**1**) abhängig zu sein, da unabhängig von der Konzentration des verfütterten Materials nur ein gewisser Teil deformyliert wird. Ein Teil des eingebauten $^{13}C_5$-$^{15}N_2$-fdC (**1**) bleibt stabil im Genom.

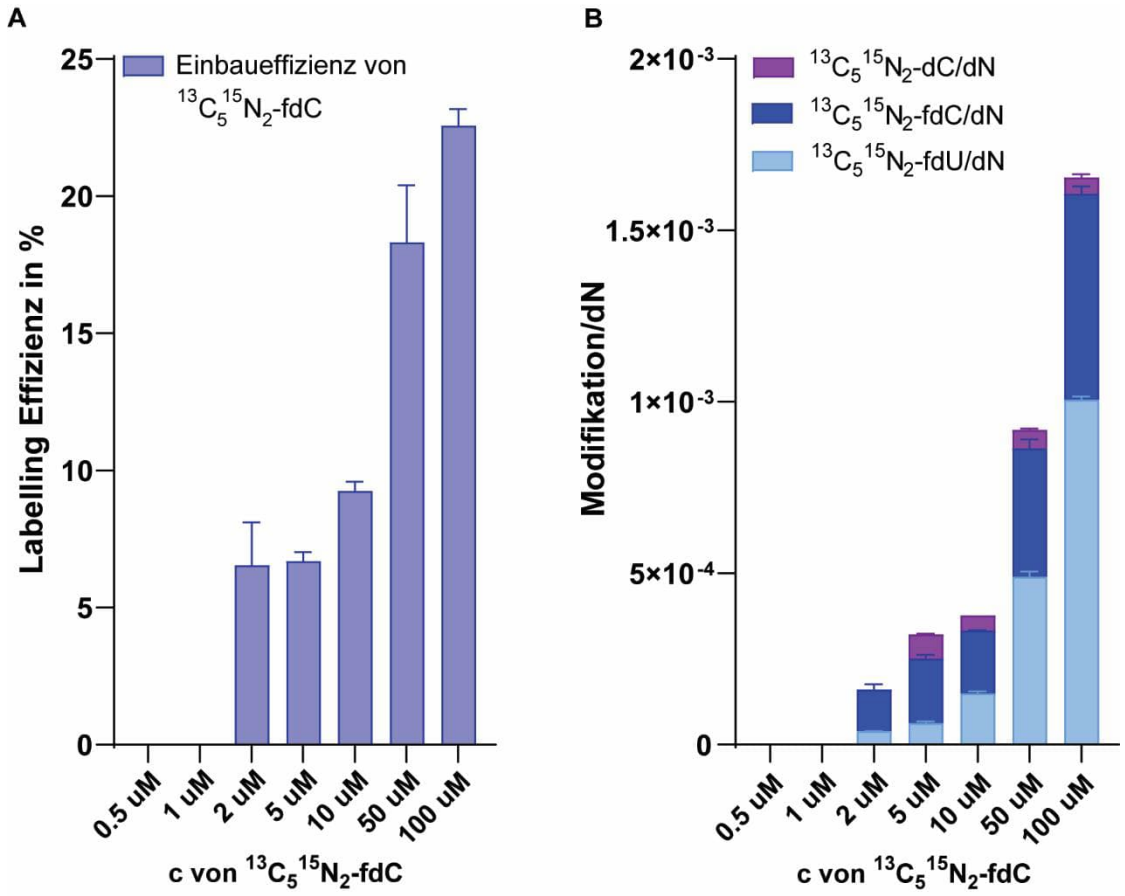


Abbildung 4.7: A Einbaueffizienz von $^{13}C_5$-$^{15}N_2$-fdC (**1**) nach Fütterung von unterschiedlichen Konzentrationen für 24 h an mESC E14 Zellen. Die Einbaueffizienz berechnete sich aus der Menge von **1** im Genom geteilt durch die Menge von genomischen fdC. **B** Absolute Level an $^{13}C_5$-$^{15}N_2$-dC/dN, $^{13}C_5$-$^{15}N_2$-fdC/dN und $^{13}C_5$-$^{15}N_2$-fdU/dN nach Verfütterung von **1** in unterschiedlichen Konzentrationen an mESC E14 Zellen nach 24 h.

Wie in **Abbildung 4.7 B** gezeigt wurde, wird nach dem Einbau von **1** ins Genom ein Teil dieser Modifikation durch direkte Demethylierung (Deformylierung) zu $^{13}C_5$-$^{15}N_2$-dC (**2**) umgewandelt. In weiteren Experimenten wurde untersucht, welchen Einfluss bestimmte Enzyme auf die Deformylierungsreaktion haben. Hierbei wurde analog zu der Studie mit K3 mES-Zellen ein Tdg-*Knockout* der E14 mES-Zellen verwendet. Die Verfütterungsexperimente an diese Zelllinie sollten Aufschluss darüber geben, wie dominant die Tdg-vermittelte BER im Vergleich zur Demethylierung durch einen direkten C-C-Bindungsbruch, in der Regulation der genomischen fdC Menge ist. Im linken Teil der **Abbildung 4.8** sind die Mengen der isotopenmarkierten Nukleoside $^{13}C_5$-$^{15}N_2$-dC (**2**), $^{13}C_5$-$^{15}N_2$-fdC (**1**) und $^{13}C_5$-$^{15}N_2$-fdU (**13**) nach der Verfütterung von **1** an E14-Wildtyp und einen E14 Tdg-KO zu sehen. Es sind keine signifikanten Unterschiede zwischen diesen Zelllinien zu erkennen. Die Menge an $^{13}C_5$-$^{15}N_2$-fdC (**1**) war sowohl im Willdtyp als auch im Tdg-KO vergleichbar, wohingegen die Menge an Deformylierungsprodukt $^{13}C_5$-$^{15}N_2$-dC (**2**) im Tdg-KO leicht erhöht war. Dies könnte durch die fehlende Aktivität der Glykosylase erklärt werden. Jedoch zeigen diese Daten eindeutig, dass

die Tdg-vermittelte Basenexzisionsreparatur von fdC keine übergeordnete Rolle in der aktiven Demethylierung spielt, da es zu keiner Akkumulation von $^{13}C_5$-$^{15}N_2$-fdC (**1**) oder einer signifikanten Erhöhung der Deformylierungsaktivität im System ohne Tdg-Aktivität kommt.

In allen hier untersuchten Zellsystemen wurden nach der Verfütterung von **1** große Mengen an $^{13}C_5$-$^{15}N_2$-fdU (**13**, bis $2x10^{-3}$/dN) im Genom detektiert. Diese hohen genomischen Level an $^{13}C_5$-$^{15}N_2$-fdU (**13**) werden als oxidative Schäden erkannt und durch Glykosylasen wie Smug1 ausgeschnitten. Durch diesen Prozess wirken hohe fdU Level toxisch für ein Zellsystem. Das Testmolekül **1** wurde an ein Smug1-KO sowie an ein CDA-*double knockout* (Cytidin Desaminase-DKO) Zellsystem verfüttert und die genomische Zusammensetzung mit dem ensprechenden Wildtyp verglichen. Hierfür wurde die embryonale Stammzellinie J1 verwendet. Die Verfütterung an Smug1-KO und CDA-DKOs sollte zeigen, welchen Einfluss der hohe Einbau von $^{13}C_5$-$^{15}N_2$-fdU (**13**) auf die Deformylierungsreaktion ausübt. Da die Smug1 Glykosylase ein Enzym ist, welches Uracil-Bausteine aus dem Genom entfernt, fehlt dieser Prozess in den untersuchten Smug1-KO Zellen, sodass das eingebaute $^{13}C_5$-$^{15}N_2$-fdU (**13**) nicht ausgeschnitten wird. Beim CDA-DKO ist eine Unterdrückung der Desaminierung von $^{13}C_5$-$^{15}N_2$-fdC (**1**) zu $^{13}C_5$-$^{15}N_2$-fdU (**13**) zu erwarten. Dadurch sollte bevorzugt das erwünschte $^{13}C_5$-$^{15}N_2$-fdC (**1**) ins Genom eingebaut werden. Das Zellmedium der Wildtyp-Zellen sowie der entsprechenden *Knockout* Zelllinien wurde mit 100 µM **1** versetzt und 24 h inkubiert. In der **Abbildung 4.8** sind die absoluten Level der isotopenmakierten Nukleoside gezeigt, die im Genom der entsprechenden Zelllinien nach Verfüttern von **1** detektiert wurden. Im linken Teil der Abbildung ist zu sehen, dass die Inaktivität von Tdg keinen signifikanten Einfluss auf die eingebaute Menge von **1** und $^{13}C_5$-$^{15}N_2$-fdU (**13**) hat. Die Level von $^{13}C_5$-$^{15}N_2$-dC (**2**) sind im Tdg-KO leicht erhöht. Dies wird im Verlauf des Kapitels noch näher diskutiert. Im rechten Teil der Abbildung sind die genomischen Level von $^{13}C_5$-$^{15}N_2$-fdC (**1**), $^{13}C_5$-$^{15}N_2$-dC (**2**) und $^{13}C_5$-$^{15}N_2$-fdU (**13**) im Smug1-KO, CDA-DKO und im entsprechenden J1-Wildtyp zu sehen. Wie zu erwarten, wurden extreme Unterschiede in den detektierten Mengen an $^{13}C_5$-$^{15}N_2$-fdU (**13**) zwischen dem Widltyp und den zwei untersuchten *Knockouts* beobachtet. Im Wildtyp wurde $^{13}C_5$-$^{15}N_2$-fdU (**13**) in Mengen von $1x10^{-3}$/dN detektiert. Im Smug1-KO lag mit einer absoluten Menge von mit $2x10^{-3}$/dN die doppelte Menge an isotopenmarkierten Uridinderivats im Vergleich zum Wildtyp im Genom vor. Durch die Inaktivität der Glykosylase kommt es zur Anreicherung von $^{13}C_5$-$^{15}N_2$-fdU (**13**) im Genom. Die Menge des eingebauten $^{13}C_5$-$^{15}N_2$-fdC (**1**) ist vergleichbar mit dem Wildtyp. Interessanterweise wurde im Smug1-KO weniger Deformylierungsprodukt als im Wildtyp detektiert. Dieser Trend wird später in diesem Kapitel näher erläutert. Die absoluten Level von $^{13}C_5$-$^{15}N_2$-fdU (**13**) im CDA-DKO unterscheiden sich stark von Wildtyp. Durch die unterdrückte Aktivität der Desaminasen kommt es zu einer stark reduzierten Desaminierung im löslichen Nukleosid Pool, wodurch die Mengen von $^{13}C_5$-$^{15}N_2$-

fdU (**13**) im Genom kleiner als 6x10^{-5}/dN sind. Die Menge des eingebauten $^{13}C_5$-$^{15}N_2$-fdC (**1**) liegt mit 2x10^{-4}/dN auch unter dem gemessenen Level von 5x10^{-4}/dN im Wildtyp. Die Rate an Deformylierung in diesen Zellsystemen wird im weiteren Verlauf des Kapitels erläutert.

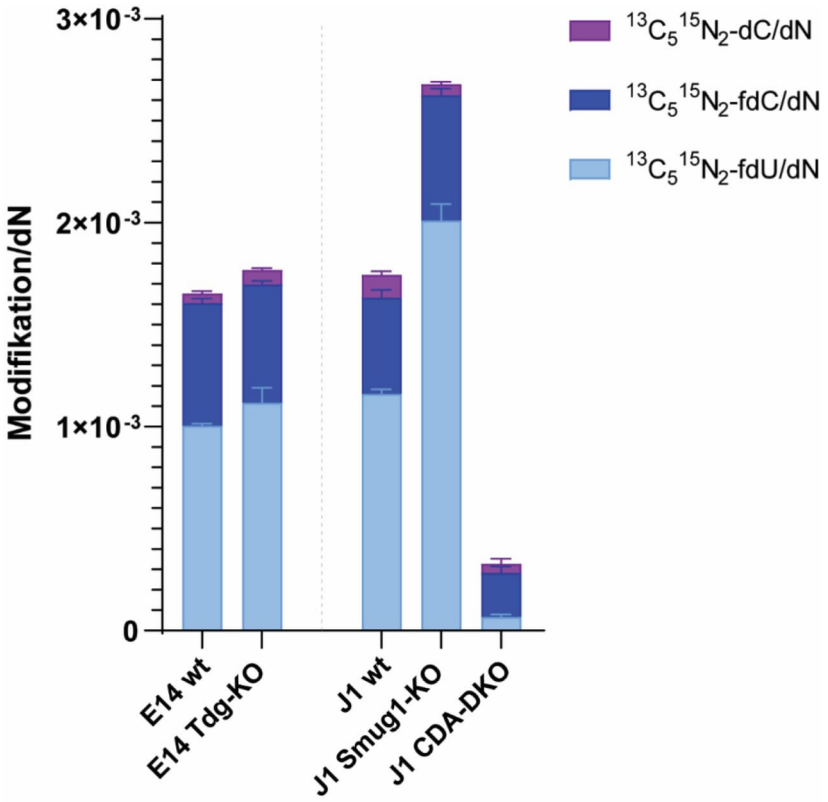


Abbildung 4.8: Absolute Level an $^{13}C_5$-$^{15}N_2$-dC/dN, $^{13}C_5$-$^{15}N_2$-fdC/dN und $^{13}C_5$-$^{15}N_2$-fdU/dN nach Verfütterung von **1** an unterschiedliche Zelllinien nach 24 h. Es wurden mESC E14-Wildtyp und mESC E14-Tdg-KO Zellen (links) und mESC J1-Wildtyp und Smug1-KO und CDA-DKO (rechts) verwendet.

Abbildung 4.9 stellt die unterschiedlichen Raten der Deformylierung in den untersuchten Zellsystemen dar. Hier ist die absolute Menge an $^{13}C_5$-$^{15}N_2$-dC (**2**), das durch die Deformylierung von $^{13}C_5$-$^{15}N_2$-fdC (**1**) entsteht, als Quotient der absoluten Menge an $^{13}C_5$-$^{15}N_2$-fdC (**1**) im Genom dargestellt. Dies wird im Folgenden als die Rate der genomischen Deformylierungsreaktion definiert. Die **Abbildung 4.9 A** zeigt, dass im Tdg-KO mehr $^{13}C_5$-$^{15}N_2$-fdC (**1**) zu $^{13}C_5$-$^{15}N_2$-dC (**2**) umgewandelt wird als im entsprechenden Wildtyp. Die Deformylierungsrate im Tdg-KO ist um den Faktor 1.5 höher als im Wildtyp (dies ist in der **Abbildung 4.9 B** gezeigt, in der die Deformylierungrate der KOs durch die Deformylierungsrate des jeweiligen Wildtyps geteilt wird. Der nummerische Wert dieses Verhältnisses wird als *Δ Deformylierung* definiert. Durch die Unterdrückung der Smug1 Aktivität kommt es zu einer verringerten Deformylierungsaktivität. Wie in **Abbildung 4.9 A/B** zu sehen ist, sinkt die Rate der Deformylierung im Smug1-KO im Vergleich zu Wildtyp deutlich. Es ist bisher nicht klar, welche Enzyme an der Deformylierung von fdC beteiligt sind. Die hier gezeigten Daten könnten einen Hinweis darauf geben, dass Smug1 oder eine ähnliche

Glykosylase an der Deformylierung beteiligt sein könnte. Da Glykosylasen, welche an BER beteiligt sind, fdC und cadC erkennen, erscheint es plausibel, dass sie, je nach Bedarf in der Zelle, auch an einer Demethylierung durch einen direkten C-C-Bindungsbruch beteiligt sein könnten. Ob Smug1 oder eine andere Glykosylase an der Deformylierung beteiligt ist, erfordert jedoch noch weitere intensive Untersuchungen. Aus der **Abbildung 4.9** ist auch ersichtlich, dass die Inaktivität der Desaminasen keinen signifikanten Einfluss auf die genomische Deformylierungsaktivität von fdC ausübt.

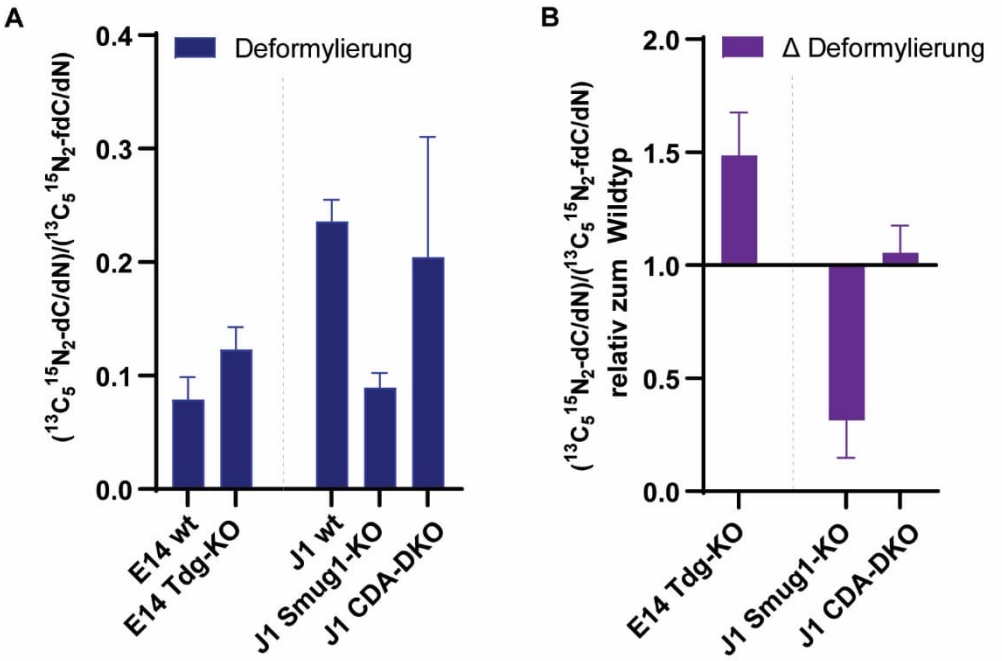


Abbildung 4.9: A Deformylierungsrate von $^{13}C_5$-$^{15}N_2$-fdC in den untersuchten Zelllinien. Die Deformylierungsrate wurde als die absolute Menge von $^{13}C_5$-$^{15}N_2$-dC/dN geteilt durch die absolute Menge von $^{13}C_5$-$^{15}N_2$-fdC/dN berechnet. **B** Zeigt die Deformylierungsrate Δ *Deformylierung* eines *Knockouts* bezogen auf die Deformylierungsrate des jeweiligen Wildtyps.

Die Verfütterungsexperimente von **1** lieferten nicht nur wichtige Daten über den Einbau von $^{13}C_5$-$^{15}N_2$-fdC (**1**) ins Genom und dessen Deformylierung zu$^{13}C_5$-$^{15}N_2$-dC (**2**). Es wurde auch gezeigt, dass das verfütterte Nukleosid $^{13}C_5$-$^{15}N_2$-fdC (**1**) im löslichen Nukelosid Pool zum großen Teil zu $^{13}C_5$-$^{15}N_2$-fdU (**13**) desaminiert und effizient ins Genom eingebaut wurde. Bei den Verfütterungsexperimenten von **1** konnte eine relative hohe Zelltodrate beobachtet werden, wobei diese im Vergleich zur Verfütterung der 2'-fluorierten Cytidinderivate deutlich höher lag. Experimente mit Tdg-KOs zeigten, dass die absoluten Mengen von $^{13}C_5$-$^{15}N_2$-fdC (**1**) im Genom unabhänig von der Tdg Aktivität sind und die hohe beobachtete Toxizität nicht mit dem Tdg-vermittelten BER Prozess zusammenhängt. Die besonders hohe beobachtete Zelltodrate nach dem Verfüttern von $^{13}C_5$-$^{15}N_2$-fdC (**1**) kann im Zusammennhang mit dem hohen Einbau von $^{13}C_5$-$^{15}N_2$-fdU (**13**) stehen. Der metabolische Weg, welcher zum Einbau von

$^{13}C_5$-$^{15}N_2$-fdU (**13**) ins Genom nach der Verfütterung von $^{13}C_5$-$^{15}N_2$-fdC (**1**) führt, ist in **Abbildung 4.10** schematisch dargestellt.

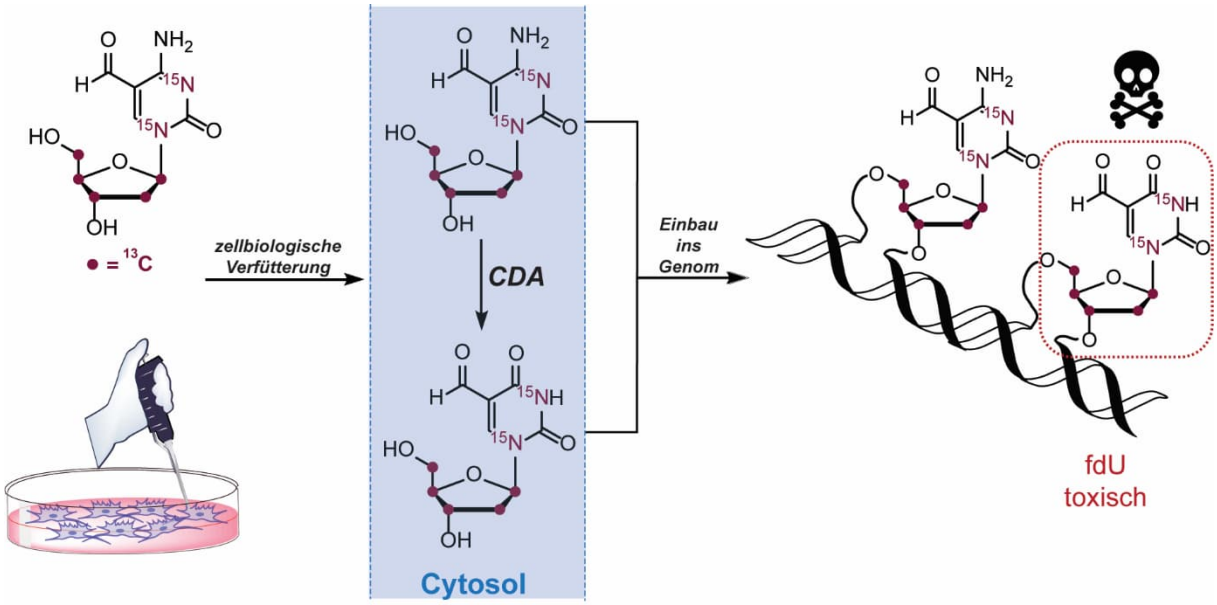


Abbildung 4.10: Experimenteller Aufbau der $^{13}C_5$-$^{15}N_2$-fdC (**1**) Verfütterung und eines dessen möglichen Metabolismuswege, Desaminierung zu $^{13}C_5$-$^{15}N_2$-fdU (**13**), im Cytosol. Anschließend können beide Modifikationen phosphoryliert und ins Genom eingebaut werden.

Aus früheren Studien ist bekannt, dass hmdU und fdU durch die Thymidin- und Thymidilat Kinase zu den ensprechenden Nukleotiden phoyphoryliert werden.[186] Analog zu hmdU, wenn fdU phosphoryliert und ins Genom eingebaut wird, wird es als geschädigte Base erkannt und triggert erhöhte Aktivität von Uracil Glykosylasen, was in Strangbrüchen resultiert.[186-187] Die erhöhte Aktivität der Smug1 Glykosylase und die daraus resultierende hohe Anzahl an schädlichen Strangbrüchen führt zu der beobachteten hohen Toxizität von fdU. Ein analoges Szenario kann für das isotopenmarkierte $^{13}C_5$-$^{15}N_2$-fdU erwartet werden. In den hier durchgeführten Verfütterungsexperimenten wurde eine hohe Zellsterberate nach Zugabe und Inkubation von $^{13}C_5$-$^{15}N_2$-fdC (**1**) ins Zellmedium beobachtet. Es wurde hierbei ein Zusammenhang zwischen Zellsterberate und absolutem Level von $^{13}C_5$-$^{15}N_2$-fdU (**13**) im Genom beobachtet**. Abbildung 4.11 A** und **B** zeigt drei unterschiedliche Zellsysteme, J1 CDA-DKO, E14 Wildtyp und J1 Smug1 KO, und die gemessene Zelltoxizität nach 24 h Verfüttern von **1** (**A**) sowie die absoluten Level von $^{13}C_5$-$^{15}N_2$-fdU (**13**) gemessen im Genom des ensprechenden Zellsystems (**B**). Es ist klar zu erkennen, dass die Toxizität proportional zur Menge des ins Genom eingebauten $^{13}C_5$-$^{15}N_2$-fdU (**13**) steigt (**Abbildung 4.11 C**). Die hohen Mengen an $^{13}C_5$-$^{15}N_2$-fdU (**13**) werden als geschädigte Basen erkannt und führen zu einer erhöhten Aktivität der Uracil Glykosylasen wie Smug1 und somit zu einer erhöhten Aktivierung der BER. Dies verursacht zahlreiche abasischen Stellen, welche zu Strangbrüchen führen können, was die hohe Toxizität von $^{13}C_5$-$^{15}N_2$-fdU (**13**) erklärt.[187] Interessanterweise wurde in

Smug1 KO auch eine hohe Sterberate von Zellen beobachtet. Die Akkumulation von von $^{13}C_5$-$^{15}N_2$-fdU im Genom bei unterdrückter BER führt also ebenso zu einer hohen Toxizität und Sterberate der untersuchten Zellsysteme.

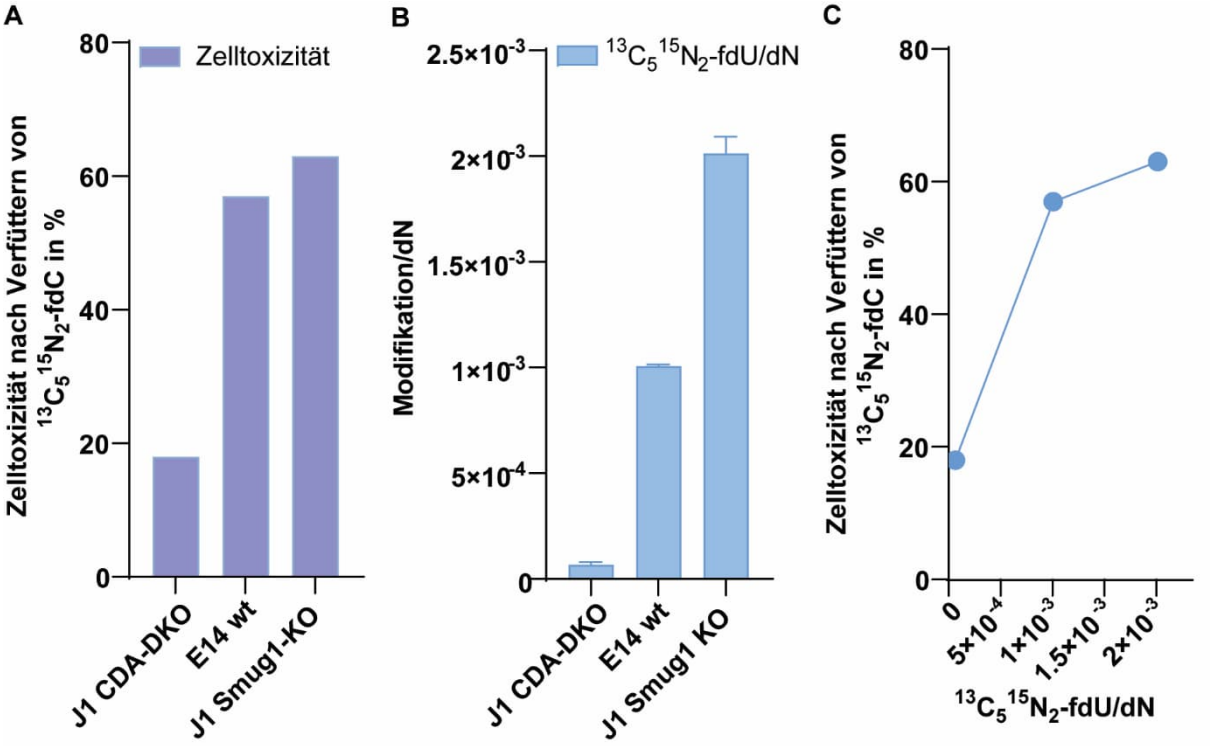


Abbildung 4.11: 1 wurde J1 CDA-DKO, E14 Wildtyp und J1 Smug1 KO gefüttert und die genomische DNA nach 24 h analysiert. **A** zeigt die Zelltoxizität bzw. die Zellsterberate nach dem durchgeführten Verfütterungsexperiment. **B** zeigt die absoluten Level an $^{13}C_5$-$^{15}N_2$-fdU (**13**) im Genom. **C** zeigt die Korrelation zwischen $^{13}C_5$-$^{15}N_2$-fdU (**13**) Menge und Zelltoxizität.

4.1.3 Zusammenfassung und Ausblick

In einer neuartigen Synthesestrategie wurde $^{13}C_5$-$^{15}N_2$-fdC (**1**) auf einem $^{13}C_5$-$^{15}N_2$-dC-freien Syntheseweg in exzellenter Reinheit hergestellt. Durch das Verfüttern von $^{13}C_5$-$^{15}N_2$-fdC (**1**) in verschiedenen Konzentrationen an unterschiedliche Zellsysteme konnte eine genomische Einbaueffizienz von bis zu 20% erreicht werden. Dies bedeutet, dass 20% aller fdCs im Genom isotopenmarkiert waren und dadurch eine ideale Grundlage zum Studieren der genomischen Deformylierung bildeten. Die Verfütterungsexperimente von $^{13}C_5$-$^{15}N_2$-fdC (**1**) an embryonale Stammzellen und somatische Zellen zeigten, dass die Deformylierungsreaktion je nach Zelltyp unterschiedlich stark ausgeprägt ist. Dies ist zu erwarten, da je nach Zelltyp auch der Grad an Methylierung von dC und damit verbunden der Grad von aktiver Demethylierung unterschiedlich stark ausgeprägt ist. Im Gegensatz zu 2'-F-fdC (**14**), ist das $^{13}C_5$-$^{15}N_2$-fdC (**1**) im chemischen Sinne seiner Reaktivität in einem DNA-Kontext mit dem natürlich vorkommenden fdC identisch. Unter Berücksichtigung der Tatsache, dass $^{13}C_5$-$^{15}N_2$-fdC (**1**) vermutlich an zufälligen Stellen im Genom eingebaut wird, ist ein analoges Deformylierungsverhalten wie für das natürliche fdC zu erwarten. Die Deformylierung ist in

einer Vielzahl von Zelltypen, Stammzellen sowie somatischen Zellen, ausgeprägt. Die Inaktivität von Tdg, welches BER von fdC iniziiert, führt zu einer leicht erhöhten Deformylierungsrate. Der Einbau von $^{13}C_5$-$^{15}N_2$-fdC (**1**) ist im Tdg-KO mit dem im Wildtyp vergleichbar, was zur Annahme führt, dass die Tdg-vermittelte BER keine übergeordnete Rolle in der aktiven Demethylierung von fdC spielt. Untersuchung der Deformylierungsaktivität in *Knockout* Systemen (Tdg-KO, Smug1-KO und CDA-DKO) hat gezeigt, dass die Inaktivität von Tdg zu einer leichten Erhöhung der Deformylierungsaktivität führt und auf der anderen Seite, der *Knockout* der Uracil Glykosylase Smug1 zu einer deutlichen Reduzierung der Deformylierungsaktivität führt. Dies könnte darauf hinweisen, dass möglicherweise diese oder eine ähnliche Glyskosylase in der direkten C-C-Bindungspaltung von fdC beteiligt ist. Hier ist jedoch zu beachten, dass es bei diesen Daten um ein technisches Triplikat eines biologischen Duplikats handelt und noch weitere Replikate und Bedingungen untersucht werden müssen, um eine genauere Aussage treffen zu können. Die Verfütterungsexperimente von $^{13}C_5$-$^{15}N_2$-fdC (**1**) haben auch gezeigt, dass es im löslichen Nukelosid Pool zu einer starken Desaminierung von $^{13}C_5$-$^{15}N_2$-fdC (**1**) zu $^{13}C_5$-$^{15}N_2$-fdU (**13**) durch Cytidin Desaminasen kommt. Das desaminierte $^{13}C_5$-$^{15}N_2$-fdU (**13**) kann auch phosphoryliert und ins Genom eingebaut werden. Die hohen detektierten genomischen $^{13}C_5$-$^{15}N_2$-fdU (**13**) Level haben sich für die Zellen als hochtoxisch erwiesen. Um ein Bild von der Deformylierungsreaktion ohne große genomische Menge an $^{13}C_5$-$^{15}N_2$-fdU (**13**) zu bekommen, wurde ein CDA-DKO verwendet. Diese Experimente haben eine vergleichbare Deformylierungsrate wie im Wildtyp gezeigt.

Als nächsten Schritt wäre es wichtig, die mit 2'-fluorierten Derivaten gezeigte Decarboxylierung, in einem ähnlichen Kontext zu untersuchen. Es ist aus der Literatur bekannt, dass cadC im löslichen Nukleosid Pool nicht von der Cytidin Desaminase erkannt und desaminiert wird. Dadurch könnten viel höhere Einbauraten von isotopenmarkierten cadC Testmolekülen erreicht werden. Aus den Erfahrungen mit den 2'-fluorierten cadC Derivaten müsste hierfür ein isotopenmarkiertes cadC Triphosphat synthetisiert werden und mit dem bekannten Nukleotid Transporter an Zellen gefüttert werden. Mit diesem Aufbau könnte es möglich sein, die genomische Decarboxylierung als Teil der aktiven Demethylierung von mdC, mit einem natürlichen System zu untersuchen und die natürlich vorkommende Decarboxylierung von cadC genauer untersuchen.

4.1.4 Projektbeiträge

Im Rahmen dieser Arbeit war ich an der Entwicklung der neuartigen Synthese von von $^{13}C_5$-$^{15}N_2$-fdC beteiligt. Die ursprüngliche Synthese wurde von *René Rahimoff* entwickelt. Die Zellkulturversuche sowie die Isolation aller DNA-Proben wurden von *Angie Kirchner* und *Ewelina Kamińska* durchgeführt und entwickelt. Ich war an der Entwicklung der

massenspektrometrischen Methoden beteiligt, in Anlehnung auf die von *Katharina Iwan* entwickelten Methoden. Der enzymatische Verdau, die Messung und Auswertung der Daten wurden von *Ewelina Kamińska* und mir vorgenommen. Die Daten wurden gemeinsam von *Ewelina Kamińska* und mir interpretiert.

4.1.5 Material und Methoden

4.1.5.1 Synthesevorschriften und Analytik

Die Synthesevorschriften und Analytik sind im Kapitel 5.2.2 zu finden.

4.1.5.2 Enzymatischer Verdau

Der enzymatische Verdau der genomischen DNA aus den Fütterungsexperimenten mit $^{13}C_5$-$^{15}N_2$-fdC erfolgte wie im Kapitel 8.1 beschriebenen Protokoll.

4.1.5.3 UHPLC-MS/MS Analyse

Die quantitative UHPLC-MS/MS-Analyse der hydrolysierten DNA-Proben erfolgte wie im Kapitel 8.1 beschrieben mit einem 1290 UHPLC-System von Agilent, welches über einen UV-Detektor verfügt und mit einem 6490 Triple-Quadrupol Massenspektrometer von Agilent verbunden war. Die Chromatographie wurde mit einer Poroshell 120 SB-C8 Säule (Agilent, 2.7 µm, 2.1 mm × 150 mm) bei 35 °C mit einem Gradienten aus Wasser und Acetonitril mit jeweils 0.0075 % (v/v) Ameisensäure durchgeführt. Der Fluss betrug 0.35 mL/min: 0 →4 min; 0 →3.5% (v/v) MeCN; 4 →6.9 min; 3.5 →5 % MeCN; 6.9 →7.2 min; 5 →80 % MeCN; 7.2 →10.5 min; 80 % MeCN; 10.5 →11.3 min; 80 →0 % MeCN; 11.3 →13 min; 0 % MeCN. Die Quellparameter waren dieselben wie im Kapitel 5.3.3 beschrieben. Die für die jeweiligen Analyten spezifischen Parameter sind im Kapitel 5.3.3 in **Tabelle 5.1** zusammengefasst.

4.1.5.4 Auswertung

Die Auswertung erfolgte wie im Kapitel 8.1 beschrieben. Die Kalibriergeraden der verschiedenen Modifikationen wurden aus bereits veröffentlichten Publikationen der Arbeitsgruppe verwendet.[133, 188] Die Kalibiriergeraden der quantifizierten Modifikationen sowie deren Gültigkeitsbereiche sind im im Kapitel 5.3.4 in **Tabelle 5.7** und **Tabelle 5.8** zusammengefasst. Die Darstellung der Daten erfolgte mit Graphpad Prism und Adobe Illustrator.

4.2 Untersuchung der Funktion und Beteiligung von cadC an der aktiven DNA-Demethylierung

Neben fdC befindet sich auch eine geringere aber signifikante Menge cadC im Genom von Neuronen und Stammzellen.[189] Seit der Entdeckung von fdC und cadC wurde für beide Nukleoside postuliert, dass sie im Genom zu dC deformylieren bzw. decarboxylieren können und so möglicherweise an der aktiven Demethylierung der DNA beteiligt sind. Für fdC wurde durch zellbiologische Verfütterungsexperimente von 2'-fluorierten sowie schwer-isotopenmarkierten Nukleosidderivaten erfolgreich gezeigt, dass fdC *in vivo* in einer Deformylierungsreaktion zu dC im genomischen Kontext demodifiziert wird und damit einen Teil der aktiven Demethylierung von mdC darstellen könnte.[135, 177] Für cadC wurde erst vor kurzem eine genomische Decarboxylierung *in vivo* gezeigt (Kapitel 3.2).[190-191] Zum Zeitpunkt der ersten drei Jahre während der Anfertigung dieser Doktorarbeit wurde eine Decarboxylierung von cadC ausschließlich *in vitro* gezeigt[93, 133, 136-137] und es war nicht bekannt, ob eine Decarboxylierung von cadC im Genom von Stammzellen vorkommt.[133, 136-137, 192] Für die Studien zur genomischen Decarboxylierung wurden als Testmoleküle analog zur Deformylierungsstudien 2'-fluorierte Nukeloside entwickelt. Für diese Verfütterungs-experimente wurde das 2'-F-cadC chemisch modifiziert um die Penetration durch die Zellmembran zu vereinfachen. Da 2'-F-cadC unter physiologischen Bedingungen zum Teil als negativ geladenes Carboxylat vorliegt und somit eine erschwerte Zellpermeabilität aufweist, wurde es für zellbiologische Experimente chemisch modifiziert und als Carbonsäureester maskiert. Diese Carbonsäureester modifizierten Probemoleküle, hier auch F-cadC *prodrugs* genannt, können viel einfacher durch die Zellmembran in das Zytosol eindringen (**Abbildung 4.12**). Die im Zytosol reichlich vorhandenen Esterasen, Enzyme die für die Hydrolyse von Carbonsäure- und Phosphorsäureestern in der Zelle zuständig sind, sollten in der Lage sein, die F-cadC *prodrugs* zu hydrolisieren und 2'-F-cadC (**15**) ins Zytosol freizusetzen.[193-197] Das 2'-F-cadC Nukleosid (**15**) kann in Anschluss von Kinasen phosphoryliert werden und in das ensprechende für den Einbau benötigte Nukleotid umgewandelt werden. Es ist aus *in vivo* Studien bekannt, dass die für die Cytidin Phosphorylierung zuständige Cytidin Monophosphat Kinase 1 (CMPK1) cadC *in vivo* nicht effizient phosphoryliert.[186] Das 2'-fluorierte Analogon könnte jedoch von einer anderen von den reichlich im Zytosol vorhandenen Kinasen erkannt und phosphoryliert werden. Falls die Phosphorylierung erfolgreich ist, kann das 2'-F-cadC Nukleotid ins Genom eingebaut werden. Nach dem Einbau ins Genom könnte eine genomische Decarboxylierung von **15** stattfinden und untersucht werden. Bei einer potenziellen Decarboxylierung von 2'-F-cadC (**15**) kann im Genom das Decarboxylierungs-produkt 2'-F-dC (**16**) detektiert werden. Die schematische Darstellung des verwendeten F-cadC *prodrug* Ansatzes kann der **Abbildung 4.12** entnommen werden.

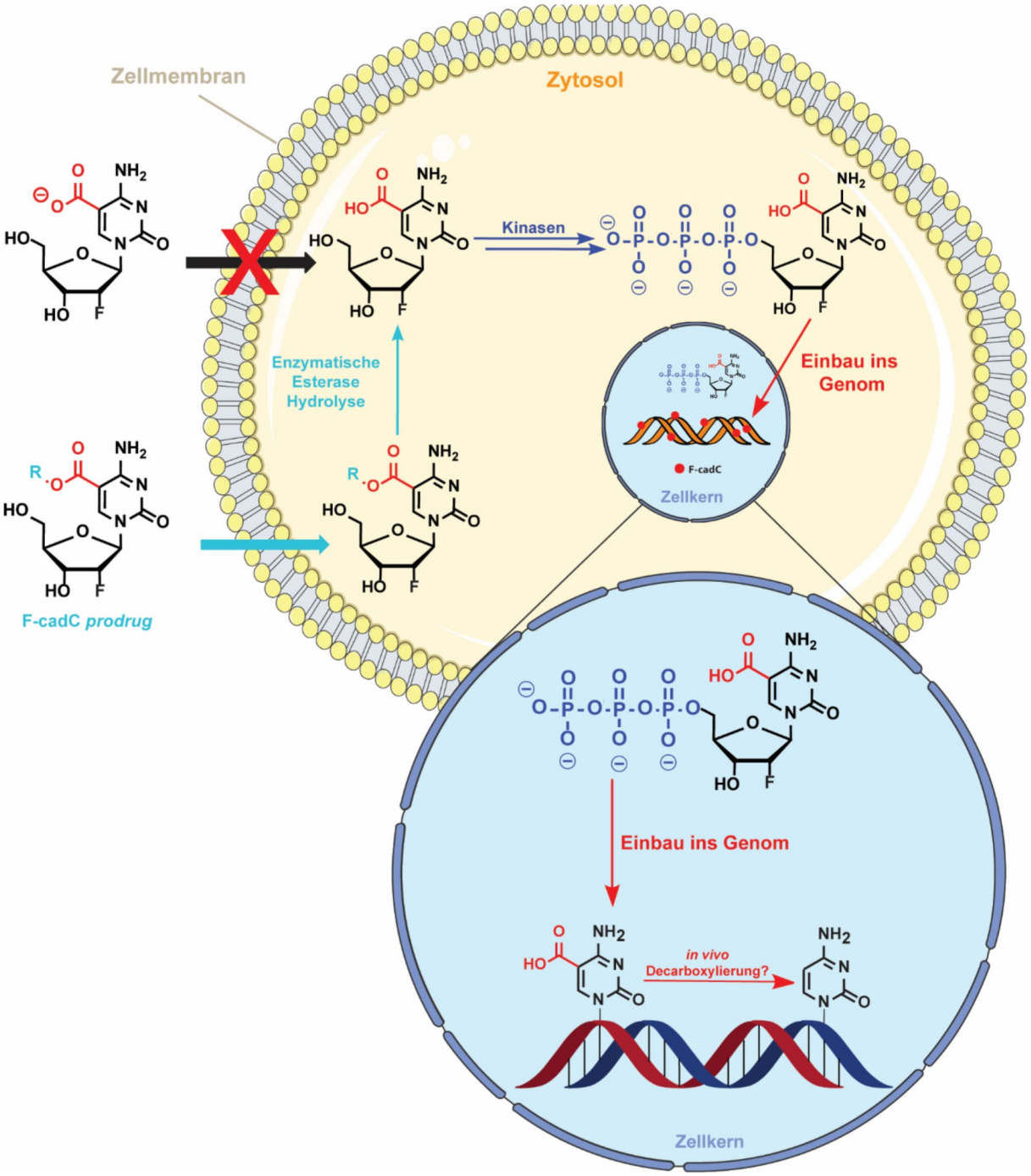
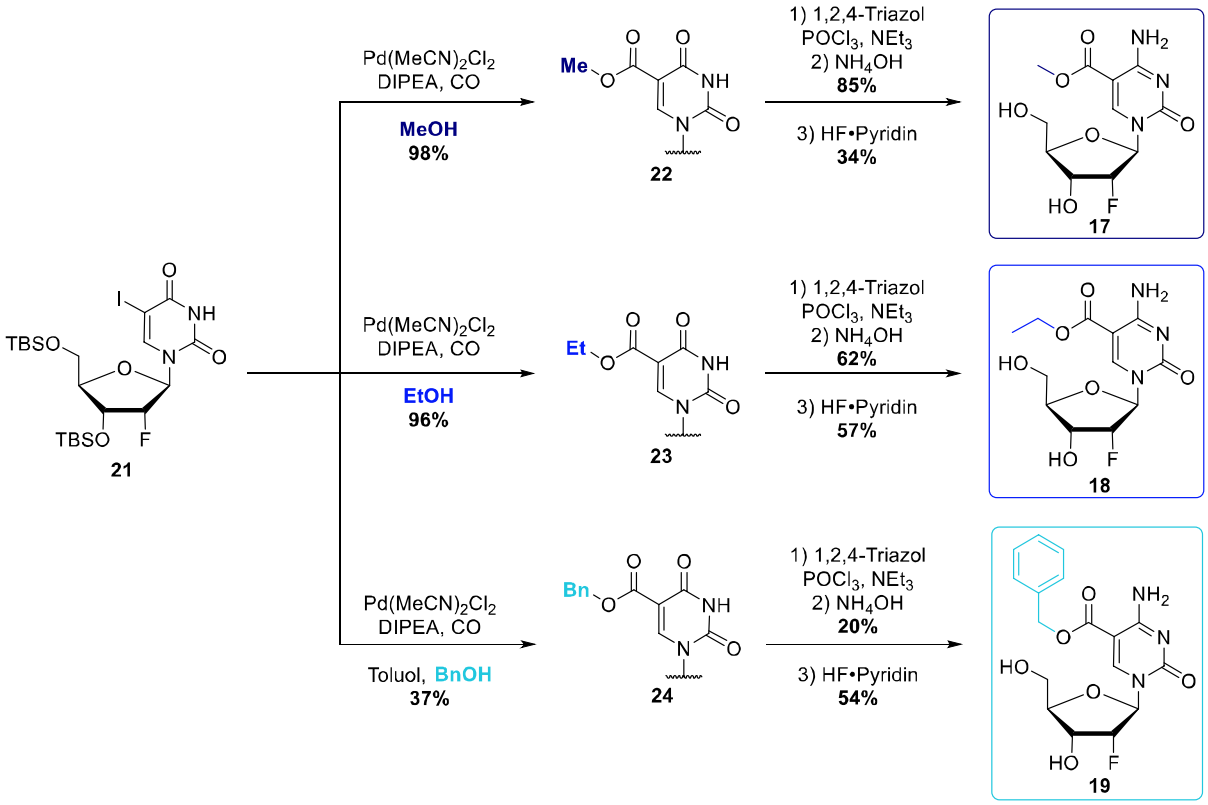


Abbildung 4.12: F-cadC *prodrug* Ansatz zur Untersuchung der genomischen Decarboxylierung von cadC *in vivo*. Hierfür werden Carbonsäureester Derivate von 2'-F-cadC eingesetzt die die Zellmembran gut durchdringen können. Im Zytosol kann die Ester Bindung der F-cadC *prodrugs* hydrolytisch von Kinasen gespalten werden und das 2'-F-cadC freisetzen. Nach der Phosphorylierung zum Nukleotid durch Kinasen und dem Einbau in die DNA kann der Prozess der Decarboxylierung von 2'-F-cadC zu 2'-F-dC untersucht werden.

4.2.1 Synthese der cadC-Prodrugs für *in vivo* Studien

Als Testmoleküle für die zellbiologischen Verfütterungsexperimente wurden drei unterschiedliche 2'-F-cadC Ester synthetisiert. Es handelt sich um das Methyl- **17**, Ethyl- **18** und Benzylcarbonsäureester **19** von 2'-F-cadC, die im Zytosol enzymatisch hydrolysierbar sein

sollten.[198] Synthetisch lassen sich diese Nukleoside in einer Funktionalisierung der Nukleobase an der C-5 Position aus dem iodiertem TBS-geschützten Uridin-Derivat, 5-Iodo-2'-F-dU (**20**) darstellen. Die Synthese dieses Vorläufers wurde von uns berichtet.[177] Alle drei Nukleosidderivate lassen sich mit einer einheitlichen Synthesestrategie darstellen. Das iodierte Desoxyuridin **21** lässt sich in einer Pd-katalysierten Carboxylierung in Anwesenheit eines Alkohols zum entsprechenden Carbonsäureester umsetzten (**Schema 4.5**). Unter diesen Bedingungen wurde **21** mit $Pd(MeCN)_2Cl_2$ und DIPEA in MeOH, EtOH oder BnOH in Toluol bei 3.5 bar CO zu 5-CO_2Me-2'-F-dU (**22**, 98%), 5-CO_2Et-2'-F-dU (**23**, 96%) und 5-CO_2Bn-2'-F-dU (**24**, 37%) umgesetzt. Die carboxylierten dU-Derivate **22**, **23** und **24** konnten mit $POCl_3$ und Triazol intermediär zu den 4-Triazolyierten Nukleosiden umgesetzt werden. Das Triazol konnte im Anschluss als gute Abgangsgruppe durch Behandlung mit wässriger NH_3-Lösung zu den ensprechenden aminierten dC-Derivaten **25** (85%), **26** (62%) und **27** (20%) umgewandelt werden. Die Abspaltung der TBS-Schutzgruppen mit HF•Pyridin lieferte nach HPL-chromatographischer Aufreinigung die Testmoleküle **17** (34%), **18** (57%) und **19** (54%) in sehr hoher Reinheit, die für die zellbiologischen Experimente erforderlich ist. Wie im Kapitel 4.1.1 beschrieben wurde, wird bei der Synthese der cadC *prodrugs* analog zu der Synthese des isotopenmarkierten fdC Derivats **1** eine Synthesestrategie gewählt, bei der das Vorkommen von 2'-F-dC vermieden wird. Durch die 2'-F-dC-freie Synthesestrategie sollten potenzielle Kontaminationen ausgeschlossen werden, die zur Verfälschung der biochemischen Experimente durch den bevorzugten Einbau von einer 2'-F-dC Verunreinigung entstehen könnten. Die Syntheseroute der F-cadC *prodrugs* **17**, **18** und **19** ausgehend von **21** ist in **Schema 4.5** dargestellt.



Schema 4.5: Syntheseschema der F-cadC *prodrugs* **17**, **18** und **19** ausgehend von 5-iodiertem 2'-F-dU (**21**), die in zellbiologischen Verfütterungsexperimenten für die Untersuchung der Decarboxylierung von cadC *in vivo* eingesetzt werden.

4.2.2 Zellbiologische Daten der Fütterungsexperimente von F-cadC

Die F-cadC *prodrugs* wurden als Testmoleküle für Verfütterungsexperimente eingesetzt um die direkte Decarboxylierung von cadC als Teil der aktiven genomischen Demethylierung zu untersuchen. Die ersten *in vivo* Studien der Decarboxylierung wurden in mausembryonalen Stammzellen durchgeführt und haben sich als ein ideales Modellsystem für diese Studien ausgezeichnet.[135] Die K3 mausemryonale Stammzellinie wurde mit einem der synthetisierten F-cadC *prodrugs* **17**, **18** oder **19** (in einer Konzentration von 350 µM im Zellmedium) versetzt und für 72 h inkubiert. Aus den Untersuchungen der Deformylierung von 2'-F-fdC (**14**) ist bekannt, dass bei diesem Prozess nach 72 h ein großer Teil des eingebauten 2'-F-fdC (**14**) zu (**2**) 2'-F-dC umgesetzt wurde.[135, 177] Falls eine Decarboxylierung von 2'-F-cadC (**15**) im Genom stattfindet, kann hier eine ähnliche oder noch höhere Rate der Demodifizierung erwartet werden. Aus chemischer Sicht ist vorstellbar, dass eine Decarboxylierung von cadC im Vergleich zur Deformylierung von fdC effizienter abläuft, da die Freisetzung von CO_2 entropisch und dadurch energetisch stark begünstigt ist. Der schematische Aufbau dieser Experimente ist in **Abbildung 4.13** dargestellt.

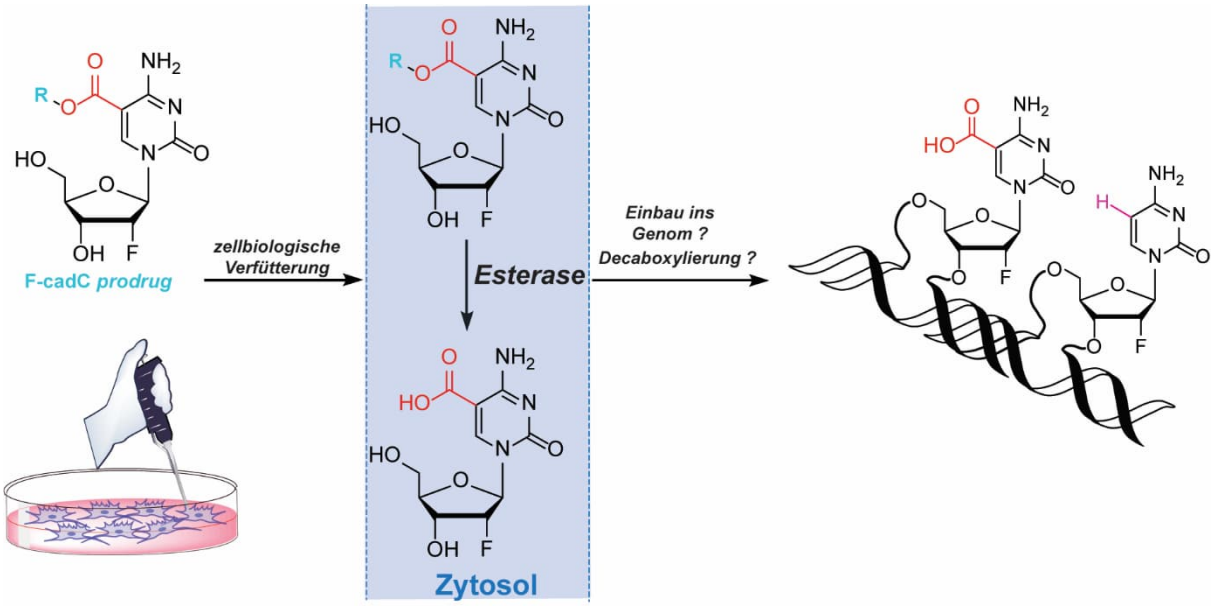
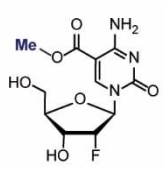


Abbildung 4.13: Experimenteller Aufbau der F-cadC *prodrug* Verfütterung an mausembryonale Stammzellen und deren enzymatische Hydrolyse zu 2'-F-cadC im Zytosol. Anschließend können die Modifikationen phosphoryliert und ins Genom eingebaut werden. Im Genom kann eine Decarboxylierung von 2'-F-cadC zu 2'-F-dC stattfinden.

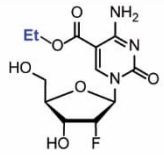
In den zellbiologischen Verfütterungexperimenten der F-cadC *prodrugs* an mESCs wurde nicht nur die genomische DNA, sondern auch das Zytosol auf das Vorkommen der 2'-fluorierten dC Derivate massenspektrometrisch analysiert. In erster Linie wurde untersucht, ob die verfütterten F-cadC *prodrugs* (2'-F-CO_2Me-dC **17**, 2'-F-CO_2Et-dC **18**, 2'-F-CO_2Bn-dC **19**) im Zytosol zu finden sind und zu 2'-F-cadC hydrolysiert wurden. Für alle verwendeten Testmoleküle wurde sowohl das entsprechende F-cadC *prodrug* als auch 2'-F-cadC (**15**) erfolgreich im Zytosol detektiert. Dies bedeutet, dass die Testmoleküle in die Zelle aufgenommen wurden und durch die Aktivität der zytosolischen Esterasen zu 2'-F-cadC (**15**) hydrolysiert wurden. Sowohl die aliphatischen Me- und Et-, als auch der Bn-Carbonsäureester wurden von Esterasen erkannt und haben 2'-F-cadC (**15**) in der Zelle freigesetzt. Eine Decarboxylierung auf der Zytosol Ebene konnte ausgeschlossen werden, da bei der Verfütterung aller hier verwendeten F-cadC *prodrugs* kein 2'-F-dC (**16**) im löslischen Nukelosidpool der Zelle detektiert wurde (**Abbildung 4.14**).

In der genomischen DNA konnten im Fall aller F-cadC *prodrugs* keine quantifizierbaren Mengen 2'-F-cadC (**15**) oder der verfütterten Ester (**17**, **18**, **19**) detektiert werden. Aus unseren später durchgeführten veröffentlichten Studien mit 2'-F-cadC Triphosphat als Testmolekül (Kapitel 3.2)[190] ist bekannt, dass die Detektion von 2'-F-cadC (**15**) im genomischen Kontext durch die extrem starke Signalunterdrückung bzw. Ionensupression erschwert ist und das ins Genom eingebaute 2'-F-cadC (**15**) nicht einfach detektiertbar ist. Wie in der von uns veröffentlichten Arbeit gezeigt wurde (Kapitel 3.2), konnten wir durch die Verfütterung eines

2'-F-cadC Triphosphats einen viel effizienteren Einbau ins Genom gewährleisten und konnten durch eine verbesserte DNA-Probenvorbereitung (Extraktion der verdauten DNA mit $CHCl_3$) die Detektion verbessern und 2'-F-cadC im Genom detektieren. Bei der Verfütterung der hier beschriebenen F-cadC *prodrugs* (**17**, **18**, **19**) konnten wir zum Zeitpunkt der Durchführung der Experimente kein 2'-F-cadC (**15**) im Genom detektieren. Jedoch wurde nach der Verfütterung aller F-cadC *prodrugs* (**17**, **18**, **19**) das Produkt der Decarboxylierung, 2'-F-dC (**16**), als auch das remethylierte 2'-F-mdC (**28**) im Genom detektiert (**Abbildung 4.14**). Diese Experimente gaben eindeutige Hinweise darauf, dass *in vivo* eine Decarboxylierung von 2'-F-cadC (**15**) auf genomischer Ebene stattfindet und dass sich das so entstandene 2'-F-dC (**15**) durch Methylierung zum 2'-F-mdC (**28**) analog zu genomischen dC verhält. Da die Detektion von 2'-F-cadC (**15**) in der DNA durch die hohe Ionensupression des Signals im genomischen Kontext erschwert ist, ist es zu diesem Zeitpunkt nicht möglich eine Einbau- oder eine Decarboxylierungseffizienz zu berechnen. Jedoch haben diese Experimente die ersten Hinweise auf eine genomische Decarboxylierung von 2'-F-cadC (**15**) gegeben und eine Basis für die Entwicklung von weiterführenden Studien gebildet. Die Ergebnisse der F-cadC *prodrug* Verfüttterung an mESCs, bei der das Zytosol und die genomische DNA auf das Vorhandensein 2'-fluorierter dC Derivate untersucht wurde, sind in **Abbildung 4.14** zusammengefasst.



	F-cadC *prodrug*	F-cadC	F-dC	F-mdC
Zytosol	✓	✓	✗	✗
Genom	✗	✗	✓	✓



	F-cadC *prodrug*	F-cadC	F-dC	F-mdC
Zytosol	✓	✓	✗	✗
Genom	✗	✗	✓	✓



	F-cadC *prodrug*	F-cadC	F-dC	F-mdC
Zytosol	✓	✓	✗	✗
Genom	✗	✗	✓	✓

Abbildung 4.14: Qualitative Ergebnisse der Verfütterungexperimente der F-cadC *prodrugs* an die K3-mES Zellen. Es wurde jeweils die zytosolische und die genomische Fraktion untersucht. In Zytosol wurde bei allen verwendeten Testmolekülen das verfütterte *prodrug* als auch 2'-F-cadC detektiert. Im Genom wurde bei allen *prodrugs* sowohl das Produkt einer Decarboxylierung, 2'-F-dC, als auch das remethylierte 2'-F-mdC detektiert.

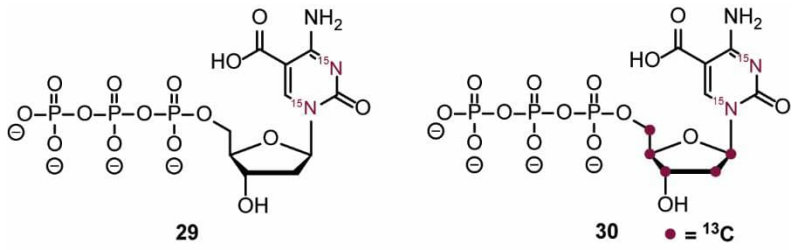
Die Detektion von 2'-F-dC (**16**) in genomischer DNA nach der Verfütterung von einem 2'-F-cadC Derivat hat die ersten Hinweise auf eine genomische Decarboxylierung als ein Teil der aktiven Demethylierung gegeben und hat den Meilenstein für die Weiterentwicklung der Testmoleküle gelegt. Für die Weiterentwicklung der Testmoleküle war zum einen ein effizienterer Einbau der 2'-F-cadC Testmoleküle als auch eine verbesserte Analyse notwendig. Wie im Kapitel 3.2 beschrieben wurde, konnte durch den Einsatz von 2'-F-cadC Triphosphat zusammen mit einem Cyclodextrin-basierten Triphosphat Transporter erfolgreich eine Decarboxylierung *in vivo* gezeigt werden. *Feng* und Mitarbeiter konnten zeitgleich zu unserer Studie in einem anderen Zellsystem (HEK- und Neuro-2a-Zellen) durch den Einsatz von 2'-F-cadC ebenso eine genomische Decarboxylierung von cadC *in vivo* zeigen.[191]

4.2.3 Zusammenfassung und Ausblick

Mit der hier gezeigten Studie wurde die genomische Decarboxylierung von cadC als ein potenzieller Teil der aktiven Demethylierung von mdC untersucht. Als Testsubstrate wurden 2'-fluorierte cadC Derivate synthetisiert und an mausembryonale Stammzellen verfüttert. Um eine möglichst hohe Zellpermeabilität zu erreichen wurde die Carbonsäurefunktionalität von cadC maskiert und es wurden drei unterschiedliche Carbonsäureester von 2'-F-cadC, Me-, Et- und Bn-Ester (**17**, **18**, **19**), getestet. Massenspektrometrische Analyse des Zytosols hat bestätigt, dass die F-cadC *prodrugs* in den löslichen Nukleosidpool gelangen konnten und hier zu 2'-F-cadC (**15**) hydrolisiert wurden. Im nächsten Schritt wurde untersucht, ob 2'-F-cadC (**15**) in die genomische DNA der mESCs eingebaut wurde und dort zu 2'-F-dC (**16**) decarboxylierte. Es konnte im Fall aller untersuchten F-cadC *prodrugs* (**17**, **18**, **19**) kein 2'-F-cadC (**15**) im Genom detektiert werden, jedoch wurde in allen Fällen das Decarboxylierungsprodukt 2'-F-dC (**16**) im Genom detektiert. Da im Zytosol kein 2'-F-dC (**16**) detektiert wurde, kann in erster Linie eine Decarboxylierung im Nukleosidpool und ein anschließender Einbau ins Genom ausgeschlossen werden. Unsere späteren Untersuchungen haben bestätigt, dass die Detektion von 2'-F-cadC (**15**) im genomischen Kontext durch eine sehr hohe Ionensupression des Signals erschwert ist. Wenn ein kleiner Teil des zytosolischen 2'-F-cadC (**15**) ins Genom eingebaut wird, kann das Signal möglicherweise so stark unterdrückt werden, dass dadurch keine quantifizierbaren Messergebnisse erzielt werden können. Das eindeutig im Genom vorhandene 2'-F-dC (**16**) sowie dessen Methylierungsfolgeprodukt 2'-F-mdC (**28**) deutet jedoch auf eine Decarboxylierung von 2'-F-cadC (**15**) im Genom. Möglicherweise ist die Einbaurate beim Verfüttern der F-cadC *prodrugs* nicht ideal, da hier mehrere enzymatische Schritte für den Einbau ins Genom ablaufen müssen. Für die Untersuchung eines dynamischen Prozesses auf DNA-Ebene ist es essenziell, dass eine ausreichende Menge des Testmoleküls in die DNA eingebaut wird um zum einen das Testmolekül selbst, zum anderen deren Metabolite detektierbar sind. Die hier diskutierten Verfütterungsexperimente der F-cadC *prodrugs* (**17**, **18**, **19**) haben uns erste Hinweise darauf gegeben, dass *in vivo* auf genomischer Ebene Decarboxylierung von 2'-F-cadC (**15**) stattfindet und dienten als wichtige Grundlage für die Verbesserung und Weiterentwicklung der Testmoleküle. Durch den Einsatz von einem 2'-F-cadC Triphosphat zusammen mt einem Cyclodextrin-basierten Nukleotid Transporter gelang es uns, wie im Kapitel 3.2 beschrieben, einen effizienteren Einbau von 2'-F-cadC zu erreichen und so eine genomische Decarboxylierung von cadC *in vivo* zu zeigen.

Weitere Untersuchungen der genomischen Decarboxylierung von cadC sollten auf den Einbau eines isotopenmarkierten Testmoleküls fokusiert werden. Die hier verwendeten 2'-fluorierten cadC *prodrugs* unterliegen im Genom nicht dem Einfluss der Tdg-vermittelten BER, da die

Halogenierung der Ribose zu einer starken Destabilisierung des während BER intermediär entstehenden Ribose-Oxocarbeniumions führt. Die 2'-fluorierten Testmoleküle sind herrvoragende Modellsysteme für die gezielte Untersuchung der direkten C-C-Bindungsbruchaktivität des cadC im Genom, jedoch handelt es sich bei den 2'-fluorierten Derivaten um artifizielle Modellsysteme. Durch den Einbau eines isotopenmarkierten Testmoleküls, das hinischtlich Reaktivität dem natürlich vorkommenden cadC identisch ist, könnte die Rate und Dynamik der genomischen Decarboxylierung von cadC in einem natürlicheren System studiert werden. Für diese Studie ideale Verbindungen sind wie im **Schema 4.6** gezeigt, cadC Triphosphate, bei denen entweder die Stickstoff Atome der Nukleobase mit dem ^{15}N Isotop markiert sind (**29**) oder neben der Nukleobase auch die Riboseeinheit des Nukleosid mit ^{13}C Atomen (**30**) versehen ist. Wie im Kapitel 4.1.2 besprochen kann eine hohe Desaminierungsrate der oxidierten mdC Derivate im löslichen Nukleosidpool für derartige Untersuchungen problematisch sein. Jedoch ist bekannt, dass die im löslichen Nukleosidpool vorkommende Cytidin Desaminase cadC nicht erkennt und nicht zum entsprechenden Desoxyuridin desaminiert. Dadurch ist der Einsatz von isotopenmarkierten cadC Derivaten vielversprechend. Durch Verfütterung von **29** oder **30** zusammen mit dem im Kapitel 3.2 beschriebenen Cyclodextrin-basierten Nukleotid Transporter könnte die genomische Decarboxylierung von cadC als Teil der aktiven Demethylierung in einem natürlicheren Testsystem genauer untersucht werden. Als nächster wichtiger Schritt wäre die Untersuchung der potenziellen enzymatischen Entität, die für eine genomische Decarboxylierung von cadC zuständig ist. Hierzu könnten *Knockout* Zellinien für Verfütterungsexperimente mit cadC Testmolekülen verwendet werden. In erster Linie könnten Proteine ausgeknockt werden, die bekannte cadC Interaktionspartner innerhalb der Zelle sind.[129]



Schema 4.6: Isotopenmarkierte Testmoleküle $^{15}N_2$-cadC-Triphosphat (**29**) und $^{13}C_5$-$^{15}N_2$-cadC-Triphosphat (**30**) für den Einsatz in zellbiologischen Verfütterungsexperimenten für die Untersuchung der genomischen Decarboxylierung von cadC als Teil der aktiven DNA-Demethylierung.

4.2.4 Projektbeiträge

Im Rahmen dieser Studie habe ich die Entwicklung und Synthese der F-cadC *prodrugs* durchgeführt. Die Zellkulturversuche sowie die Isolation aller DNA-Proben und löslichem Nukleosidpool wurden von *Angie Kirchner* und *Ewelina Kamińska* durchgeführt. Die Entwicklung der massenspektrometrischen Methoden, der enzymatische Verdau, die Messung und Auswertung der Daten wurden von *Katharina Iwan* und mir durchgeführt. Die Daten wurden gemeinsam von *Katharina Iwan* und mir interpretiert.

4.2.5 Material und Methoden

4.2.5.1 Synthesevorschriften und Analytik

Die Synthesevorschriften und Analytik sind im Kapitel 5.2.4 zu finden.

4.2.5.2 Enzymatischer Verdau

Der enzymatische Verdau der genomischen DNA und die Isolation des löslichen Nukleosidpools aus den Fütterungsexperimenten mit F-cadC *prodrugs* erfolgte wie im Kapitel 8.1 beschriebenen Protokoll.

4.2.5.3 UHPLC-MS/MS Analyse

Die quantitative UHPLC-MS/MS-Analyse der hydrolysierten DNA-Proben erfolgte wie im Kapitel 8.2 beschrieben mit einem 1290 UHPLC-System von Agilent, welches über einen UV-Detektor verfügt und mit einem 6490 Triple-Quadrupol Massenspektrometer von Agilent verbunden war. Die Chromatographie wurde mit einer Poroshell 120 SB-C8 Säule (Agilent, 2.7 µm, 2.1 mm × 150 mm) bei 35 °C mit einem Gradienten aus Wasser und Acetonitril mit jeweils 0.0075 % (v/v) Ameisensäure durchgeführt. Der Fluss betrug 0.35 mL/min: 0 →4 min; 0 →3.5% (v/v) MeCN; 4 →6.9 min; 3.5 →5 % MeCN; 6.9 →7.2 min; 5 →80 % MeCN; 7.2 →10.5 min; 80 % MeCN; 10.5 →11.3 min; 80 →0 % MeCN; 11.3 →13 min; 0 % MeCN. Die Quellparameter waren dieselben wie unter 8.2 beschrieben. Die für die jeweiligen Analyten spezifischen Parameter sind im Kapitel 5.3.3 in **Tabelle 5.2** zusammengefasst.

4.2.5.4 Auswertung

Die Auswertung erfolgte wie im Kapitel 8.1 beschrieben. Die Kalibriergeraden der verschiedenen Modifikationen wurden aus bereits veröffentlichten Publikationen der Arbeitsgruppe verwendet.[133, 188] Die Kalibiriergeraden der quantifizierten Modifikationen sowie

deren Gültigkeitsbereiche sind im Kapitel 5.3.4 in **Tabelle 5.7** bis **Tabelle 5.9** zusammengefasst. Die Darstellung der Daten erfolgte mit Graphpad Prism und Adobe Illustrator.

4.3 Einsatz von Nukleosid-Analoga für eine Struktur-Aktivitäts-Studie der direkten DNA-Demethylierung

Durch den Einsatz von 2'-fluorierten und isotopenmarkierten Nukleosiden konnte *in vivo* gezeigt werden, dass fdC und cadC auf genomischer Ebene zu dC deformyliert[135, 177] bzw. decarboxyliert[190-191]. Dieser direkte C-C-Bindungsbruchmechanismus könnte einen Teil der aktiven Demethylierung von mdC im Rahmen der epigenetischen Regulierung darstellen. Es wurde in *in vitro* als auch *in vivo* Studien gezeigt, dass eine genomische Deformylierung von fdC bzw. Decarboxylierung von cadC intermediär über einen Angriff eines Nukleophils an der C-6 Position des Cytosins durchläuft.[93, 133, 199] Um diesen Prozess mehr verstehen zu können und herauszufinden von welchen Faktoren eine Deformylierung/Decarboxylierung abhängig ist und unter welchen Bedingungen sie stattfindet, wurden fdC und cadC Analoga entwickelt und in zellbiologischen Verfütterungsexperimenten eingesetzt. Mit diesen Experimenten wollten wir verstehen, ob ein C-C-Bindungsbruch, der in einer Deformylierungs- und Decarboxylierungsreaktion stattfindet, spezifisch nur für eine Aldehyd- und Carbonsäure-funktionalität passiert oder ob auch andere Funktionalitäten auf diese Art demodifiziert werden. Durch den Einsatz analoger Strukturen sollte eine Struktur-Aktivitäts-Beziehungsstudie duchgeführt werden, die Informationen über die essenziellen strukturellen Teile eines Moleküls für einen bestimmten Prozess liefert. Falls die Deformylierung und Decarboxylierung enzymatische Prozesse sind, könnte das entsprechende Enzym ein chemisch ähnliches Analogon erkennen und mit der Funktionalität interagieren. Falls eine Interaktion stattfinden würde, aber die Reaktivität eines Analogons keinen C-X Bindungsbruch erlaubt, könnten auf diese Weise enzymatische Interaktionspartner gefunden werden. Die Testmoleküle werden an mESCs verfüttert und können ins Genom der Zellen eingebaut werden. Falls ein deformylierungs- oder decarboxylierungsartiger C-X Bindungsbruch passiert, sollte das entsprechende Cytidin entstehen und im Genom detektiertbar sein. Der schematische Aufbau der Studie ist in **Abbildung 4.15** zu sehen.

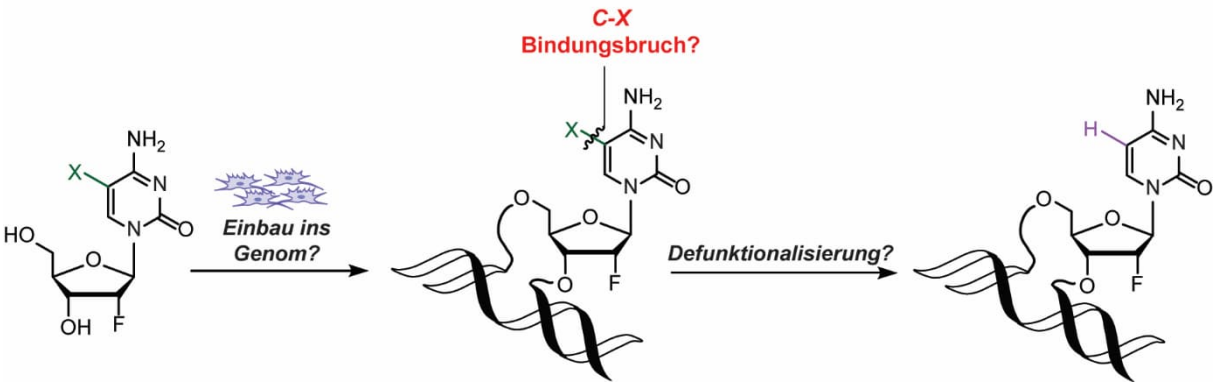
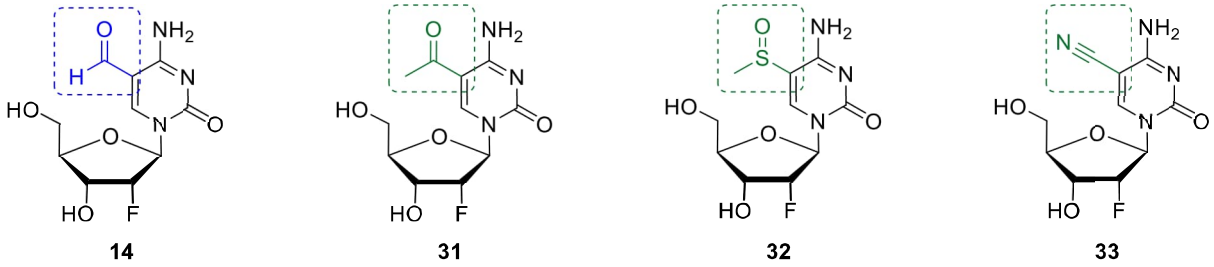


Abbildung 4.15: Experimenteller Aufbau der Verfütterungsstudien von fdC und cadC Analoga an mESCs.

4.3.1 Entwicklung und Synthese der fdC Analoga

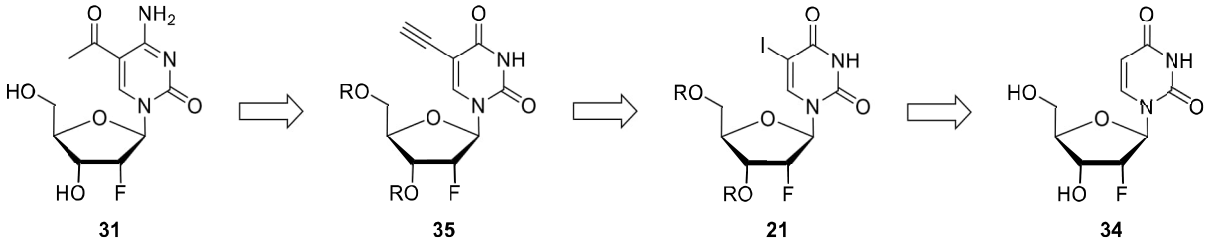
Um eine Struktur-Aktivitätsstudie durchzuführen, wurden 2'-F-dC Derivate entwickelt, die ähnliche stereoelektronische Eigenschaften und Reaktivität wie die Aldehyd Funktionalität von 2'-F-fdC besitzen, jedoch eine andere funktionelle Gruppe tragen. Aus chemischer Sicht besitzen andere Carbonylfunktionalitäten oder Systeme mit einem vergleichbaren elektrophilen Zentrum eine zum Aldehyd analoge Reaktivität. Ein Aldehyd stellt ein Elektrophil dar mit dem Carbonylkohlenstoff als elektrophiles Zentrum, das in einem 1,2-Angriff von einem Nukleophil angegriffen werden kann. Eingebettet im aromatischen System des Cytidins, stellt die Aldehydgruppe von fdC auch eine Möglichkeit zum nukleophilen Michael-artigen 1,4-Angriff, wie auch experimentell für fdC und cadC bestätigt wurde, dar.[93, 133, 199] Eine ähnliche Reaktivität kann von einem Keton-, Sulfoxid- und Nitril-substituierten 2'-F-dC erwartet werden. Diese funktionellen Gruppen sind einem Aldehyd in ihren elektronischen Eigenschaften als auch in ihrer Reaktivität ähnlich. Für zellbiologische Verfüterungsexperimente wurden folgende im **Schema 4.7** abgebildeten 2'-F-fdC Analoga entwicklet: 2'-F-COMe-dC (**31**), 2'-F-SOMe-dC (**32**), 2'-F-CN-dC (**33**).



Schema 4.7: Chemische Analoga von 2'-F-fdC, die bei zellbiologischen Verfütterungsexperimenten eingesetzt werden, um den Mechanismus der Deformylierung von fdC im Genom näher zu untersuchen.

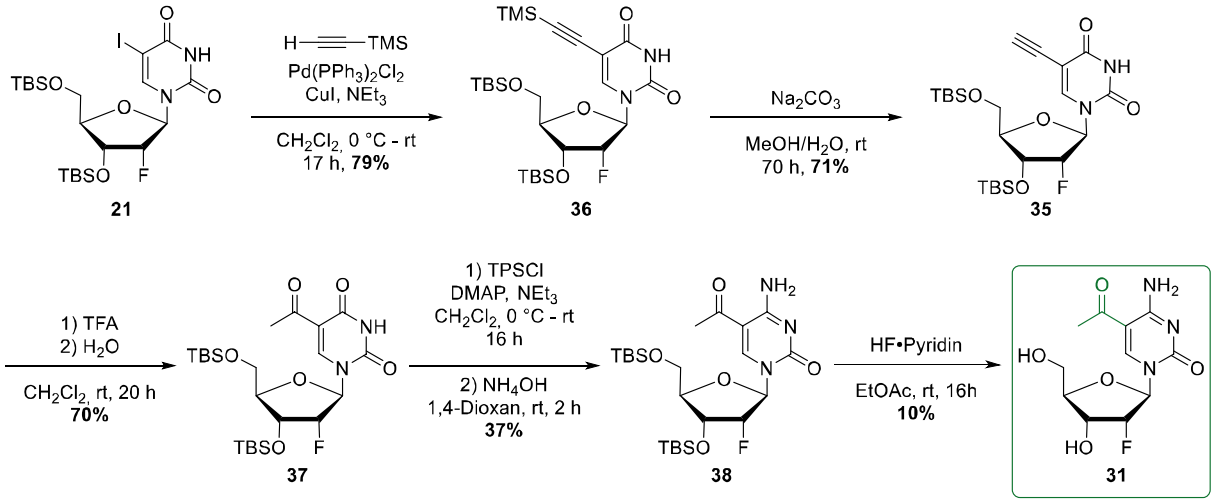
4.3.1.1 Synthese von 2'-F-fdC Analogon: 2'-F-COMe-dC (31)

Das Testmolekül 2'-F-COMe-dC (**31**) lässt sich retrosynthetisch aus kommerziell erwerblichem 2'-F-dU (**34**) darstellen. Essenziell für die Funktionalisierung ist die Aktiviterung der C-5-Position der Base, die durch eine elektrophile Iodierung (**21**) erreicht werden kann. Die Einführung der zwei Kohlenstoffatome des Carbonyls kann in Form einer Ethinylgruppe ablaufen (**35**), die durch eine Pd-katalysierte Sonogashira Kreuzkupplung erfolgt. Ein so eingeführtes terminals Alkin stellt ein maskiertes Carbonyl dar und lässt sich durch Hydratisierung in das Methylketon **31** umwandeln. Eine retrosynthetische Analyse von **31** ist im **Schema 4.8** dargestellt.



Schema 4.8: Retrosynthetische Analyse von Testmolekül 2'-F-COMe-dC (**31**).

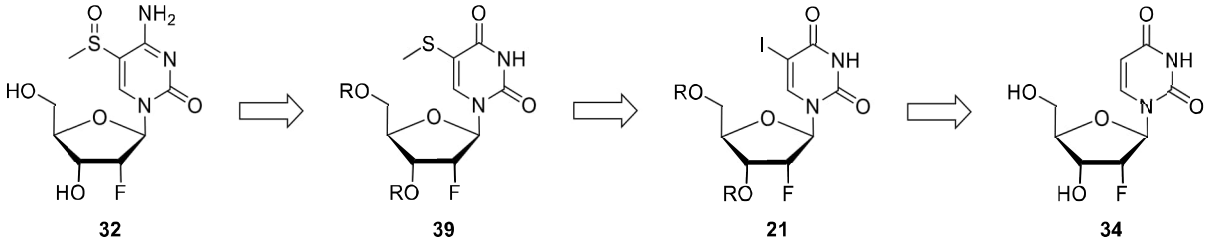
In der Synthesestrategie von **31** wurden TBS-Gruppen für die Schützung der Ribose OH-Gruppen verwendet um, zum einen die Löslichkeit der Intermediate in organischen Lösemitteln zu erhöhen und zum anderen selektiv eine Funktionalisierung an der Nukleobase vornehmen zu können. Das TBS-geschützten 2'-F-5-I-dU (**21**) wurde wie im Kapitel 3.1 beschrieben darsgestellt und stellte die ideale Vorstufe für eine Pd-katalysierte C-C-Bindungsknüpfungsreaktion dar. **21** wurde in einer Sonogashira-Hagihara Kreuzkupplung mit $Pd(PPh_3)_2Cl_2$, CuI, NEt_3 und TMS-Acetylen zu **36** in einer Ausbeute von 79% umgesetzt. Eine milde Abspaltung der Alkin-TMS Schutzgruppe durch Behandlung Na_2CO_3 lieferte das 5-Acetylen-Uridinderivat **35**. Das terminale Alkin **35** konnte durch eine Hydratisierung mit Trifluoressigsäure und Wasser zum ensprechenden Carbonyl-substituierten Nukleosid **37** umgewandelt werden. Eine Aktivierung der C-4 Position des Uridins mit Triisopropyl-sulfonylchlorid und DMAP mit einer darauffolgenden nukleophilen Substitution mit wässrigem Ammoniak führte zum carbonyl-funktionalisierten Cytidinderivat **38**. Eine finale Entschützung der TBS-Gruppen mit HF•Pyridin lieferte nach einer HPLC Aufreinigung 2'-F-5-COMe-dC (**31**) in exzellenter Reihheit. Das Syntheseschema zum Testmolekül **31** ist im **Schema 4.9** dargestellt.



Schema 4.9: Syntheseschema des Methylketon-substituierten 2'-F-dC Derivats **31** ausgehend von **21**.

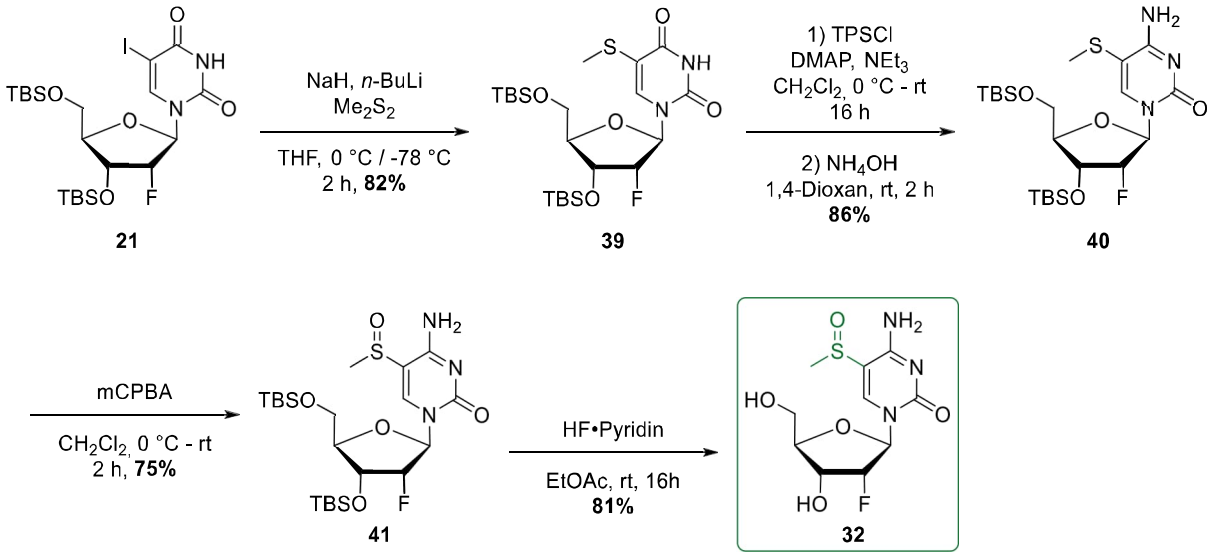
4.3.1.2 Synthese von 2'-F-fdC Analogon: 2'-F-SOMe-dC (32)

Das Sulfoxid-Analogon von 2'-F-fdC **32**, das eine C-S Verknüpfung an der Nukleobase besitzt, kann ausgehend von kommerziell erwerblichem 2'-F-dU (**34**) dargestellt werden. Die Schlüsselschritte dieser Synthesetrategie beinhalten eine 5-Funktionalisierung der Nukleobase, eine Einführung der Methylsulfonyl Funktionalität durch eine C-S-Bindungsknüpfungsreaktion und eine kontrollierte Oxidation des Schwefel Atoms zum entsprechenden Sulfoxid. Da Alkylsulfoxide eine Prädisposition zur Oxidation zu Sulfonen als auch zur C-H Deprotonierung besitzen, sollte die Oxidation zum Sulfoxid an einem späten Zeitpunkt der Synthese erfolgen. Aus diesen Überlegungen ergibt sich folgende retrosynthetische Strategie (**Schema 4.10**): eine 5-Funktionalisierung der Nukleobase zum iodiertem Nukleosid **21**, Einführung des Sulfans (**39**) gefolgt von einer Aminierung des Uridins. Im Anschluss erfolgt die Oxidation des Schwefels zum Sulfoxid **32**.



Schema 4.10: Retrosynthetische Analyse von Testmolekül 2'-F-SOMe-dC (**32**).

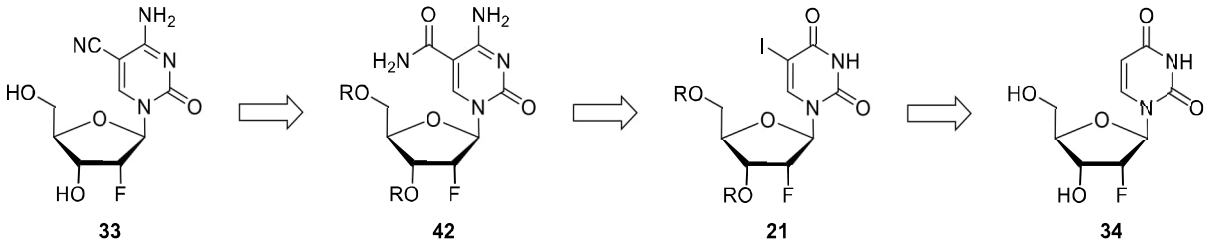
Ausgehend von kommerziell erhältlichem 2'-F-dU wurde wie im Kapitel 3.1 beschrieben, das 5-iodierte 2'-F-dU (**21**) dargestellt. **21** wurde mit NaH intermediär am *N*-3 als Na-Salz geschützt und konnte mit *n*-BuLi durch einen Iod-Lithium Austausch in ein lithiiertes C-5-Nukleophil umgewandelt werden. Nach Zugabe von Me_2S_2 fungiert das C-5-lithiierte Nukleosid als C-Nukleophil und greift am Schwefel des Me_2S_2 an. So ließ sich **21** *one pot* funktionalisieren und führte in 82% zum sulfinyliertem Uridinderivat **39**. Eine Aminierung des C-4 in zwei Schritten mit Triisopropylsulfonylchlorid und DMAP, mit anschließender Aminolyse mit Ammoniak ergab das Cytidinderivat **40**. Eine milde Oxidation des Thiol Schwefels mit *m*CPBA (**41**) und eine Abspaltung der TBS-Gruppen mit HF•Pyridin lieferte nach HPLC Aufreinigung das Sulfoxid-funktionalisierte Testmolekül **32** in exzellenter Reinheit. Das Syntheseschema zur Darstellung von **32** ist im **Schema 4.11** abgebildet.



Schema 4.11: Syntheseschema des Sulfoxid-substituierten 2'-F-dC Derivats **32** ausgehend von **21**.

4.3.1.3 Synthese von 2'-F-fdC Analogon: 2'-F-CN-dC (33)

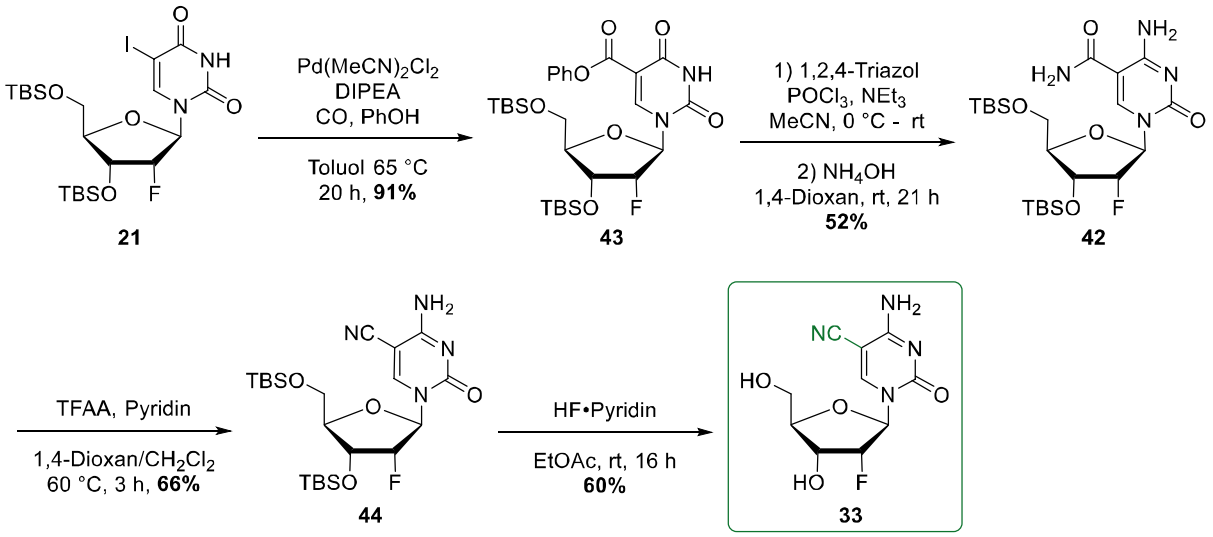
Bei der Synthese von 5-Cyano substituiertem 2'-F-dC (**33**) ist die Einführung der Cyano Gruppe der entscheidende Schritt. Die Cyano-Gruppe lässt sich retrosynthetisch durch Dehydratisierung eines Carbonsäureamids darstellen. Die Einführung des Carbonsäureamids (**42**) kann ausgehend von einem halogeniertem Nukleosid (**21**) über eine Pd-katalysierte C-C-Kreuzkupplung mit einer anschließenden Aminolyse erfolgen. Die Retrosynthetische Analyse des 5-Cyano 2'-F-dC (**33**) ist in **Schema 4.12** abgebildet.



Schema 4.12: Retrosynthetische Analyse von Testmolekül 2'-F-CN-dC (**33**).

Analog zu den in vorherigen Abschnitten beschriebenen Synthesewegen von Methylketon- (**31**) und Methylsulfoxid-substituiertem 2'-F-dC (**32**) wurde das Cyanoderivat **33** ausgehend von 5-iodiertem 2'-F-dU (**21**) hergestellt. In einer carbonylativen C-C-Kreuzkupplungsreaktion katalyisert duch $Pd(MeCN)_2Cl_2$ in Anwesenheit von DIPEA und PhOH in einer CO Atmosphäre wurde **21** in das 5-Phenoxycarboxyl 2'-F-dC (**43**) in sehr guten Ausbeuten (91%) umgewandelt. Eine Aktivierung des C-4 Kohlenstoffs der Nukleobase mit 1,2,4-Triazol und $POCl_3$ führte zum C-4 triazolyiertem Nukleosid, welches durch Zugabe von wässrigem NH_3

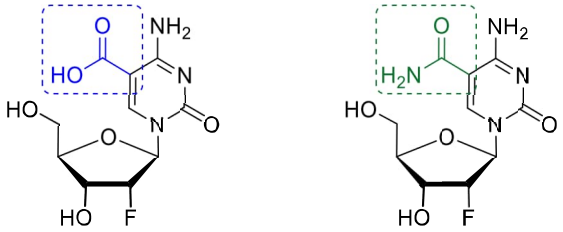
zum ensprechenden Cytidinderivat **42** führte. Unter diesen Bedingungen wurde gleichzeitig das Phenoxycarboxyl aminolysiert und zum Carbonsäureamid **42** umgewandelt. Durch dehydratisierende Reaktionsbedingungen mit Triflouressigsäureanhydrid wurde das Carbonsäureamid **42** in das 5-Cyano substituierte Nukleosid **44** überführt. Die Dehydratisierung von Carbonsäureamid **42** zum Nitril **44** konnte in einer isolierten Ausbeute von 66% erzielt werden. Eine Abspaltung der TBS-Schutzgruppen mit HF•Pyridin lieferte nach HPLC Aufreinigung das freie Nukleosid 5-CN-2'-F-dC (**33**). Der entsprechende Syntheseweg zur Darstellung des Cyanoanalogon des 2'-F-fdC (**33**) ist im **Schema 4.13** abgebidlet.



Schema 4.13: Syntheseschema des Nitril-substituierten 2'-F-dC Derivats **33** ausgehend von **21**.

4.3.2 Entwicklung und Synthese von cadC Analoga

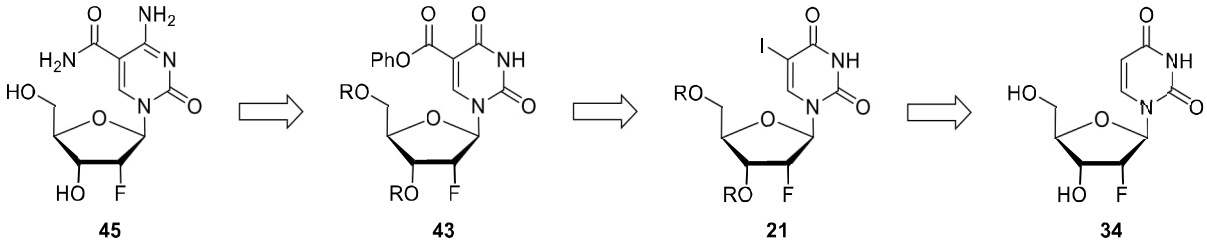
Für die Struktur-Aktivitäts-Beziehungsstudie von cadC wurde ein Nukleosidderivat von 2'-F-cadC entwickelt, das eine funktionelle Gruppe besitzt, die chemisch zu einer Carbonsäure analog ist, jedoch eine andere Reaktivität besitzt. Carbonsäurederivate neigen zu einer entropisch begünstigten Abspaltung von CO_2. Ein Carbonsäureamid besitzt eine chemisch und elektronisch analoge Struktur, kann jedoch nicht spontan analog zu einer Decarboxylierung defunktionalisiert werden. Aus diesem Grund wurde ein Carbonsäureamid funktionalisiertes 2'-F-dC **45** synthetisiert und in zellbiologischen Verfütterungsexperimenten als 2'-F-cadC Analogon eingesetzt (**Schema 4.14**)



Schema 4.14: Chemisches Analogon von 2'-F-cadC **45**, das bei zellbiologischen Verfütterungsexperimenten eingesetzt wird, um den Mechanismus der Decarboxylierung von cadC im Genom näher zu untersuchen.

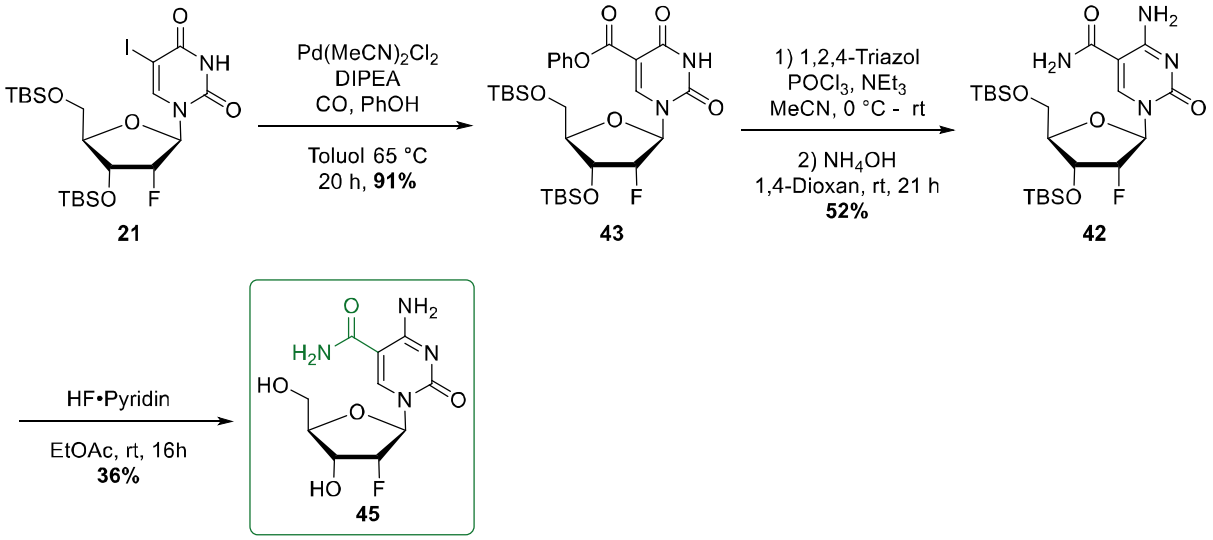
4.3.2.1 Synthese von 2'-F-cadC Analogon: 2'-F-CONH$_2$-dC (45)

Die Synthesestrategie des 5-CONH$_2$-2'-F-dC (**45**) verläuft analog zum im letzten Abschnitt beschriebenen Cyano substituierten Derivat **33**. Die Einführung der Carbonsäureamid Funktionalität erfolgt über eine Aminolyse eines Phenoxycarboxyl-substituierten Uridin Derivats **43**. Der Phenoxycarbonsäureester kann über eine carbonylative Pd-katalysierte Kreuzkupplung dargestellt werden (**Schema 4.15**)



Schema 4.15: Retrosynthetische Analyse von Testmolekül 2'-F-CONH$_2$-dC (**45**).

Die Synthese von **45** wurde analog zur Synthese von 5-CN-2'-F-dC (**33**, **Schema 4.13**) durchgeführt. Das TBS-geschützte Carbonsäureamid funktionalisierte Nukleosid **42** lieferte nach Reaktion mit HF•Pyridin nach einer HPLC Aufreinigung das 2'-F-cadC Analogon **45** wie dem **Schema 4.16** zu entnehmen ist.



Schema 4.16: Syntheseschema des Amid-substituierten 2'-F-dC Derivats **45** ausgehend von **21**.

4.3.3 Zellbiologische Fütterungsdaten

Die 2'-F-fdC Analoga 2'-F-COMe-dC (**31**), 2'-F-SOMe-dC (**32**), 2'-F-CN-dC (**33**) sowie das 2'-F-cadC Analogon 2'-F-$CONH_2$-dC (**45**) wurden in zellbiologischen Verfütterungsexperimenten eingesetzt. Es wurde untersucht, ob die in das Genom der untersuchten Zellen eingebaut werden und dort ein C-X-Bindungsbruch stattfindet. Hierzu wurden die Testmoleküle in einer Konzentration von 350 µM im Zellmedium an die Neuro-2a Zelllinie (N2a) verfüttert und für 72 h inkubiert. Im Anschluss wurde die genomische DNA sowie das Zytosol isoliert und massenspektrometrisch untersucht. Im löslichen Nukleosidpool sollte zunächst geprüft werden, ob die verfütterten Testmoleküle in die Zelle aufgenommen wurden und im zytosolischen Medium stabil sind. In der genomischen DNA, sollte mittels Massenspektrometrie untersucht werden, ob die jeweiligen Testmoleküle ins Genom eingebaut wurden und ob sie zum 2'-F-dC metabolisiert wurden. Der experimentelle Aufbau dieser Studie ist in **Abbildung 4.16** abgebildet.

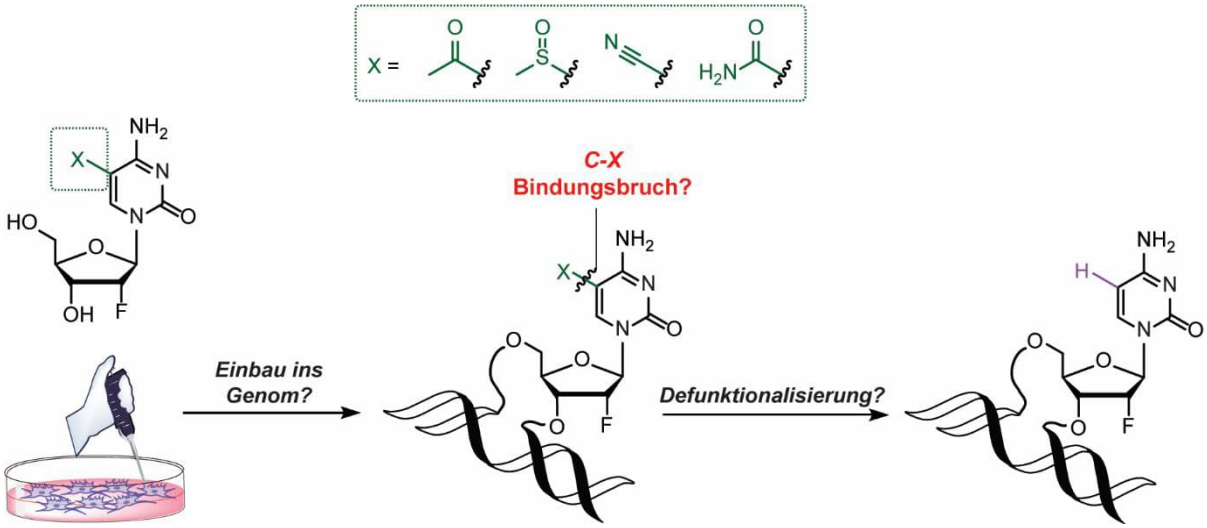


Abbildung 4.16: Experimenteller Aufbau der Verfütterungsexperimente der 2'-F-fdC und 2'-F-cadC Analoga.

Die Ergebnisse der Verfütterungsstudien der Nukleosidanaloga sind in der **Abbildung 4.17** bildlich dargestellt. Die vier verschiedene Testmoleküle 2'-F-COMe-dC (**31**), 2'-F-SOMe-dC (**32**), 2'-F-CN-dC (**33**) und 2'-F-$CONH_2$-dC (**45**) konnten alle im Zytosol der untersuchten Zellen mittels Massenspektrometrie nachgewiesen werden. Dies deutet darauf hin, dass diese Nukleoside in die Zelle aufgenommen werden können. Es wurden keine anderen 2'-fluorierten Nukleoside im Zytosol detektiert. Insbesondere war auch kein 2'-F-dC (**16**) als Abbauprodukt einer möglichen Defunktionalisierungsreaktion nachweisbar. Insgesamt lässt dies daraus schließen, dass die untersuchten Testmoleküle unter physiologischen Bedingungen stabil sind. Bei der Analyse der genomischen DNA hingegen zeigte sich jedoch, dass die Testmoleküle **31**, **32**, **33** und **45** bei der verwendeten Zelllinie nicht ins Genom eingebaut wurden, da weder die verfütterten Testmoleküle **31**, **32**, **33** und **45**, noch deren potentielles Defunktionalisierungsprodukt 2'-F-dC (**16**) in der genomischen DNA detektiert wurden. Interessanterweise wurde im Fall des 2'-F-SOMe-dC (**32**) 2'-fluoriertes dC (**16**) in der DNA detektiert. Der Einbau des 2'-F-SOMe-dC (**32**) konnte jedoch nicht direkt nachgewiesen werden. Es findet also eine Defunktionalisierung dieses Sulfoxid-derivatisierten Nukleosids unter physiologischen Bedingungen in der Zelle statt. Obwohl **32** nicht direkt in der DNA detektiert wurde, legt das Nichtvorhandensein des 2'-F-dC (**16**) im Zyotsol nahe, dass dieser C-S-Bindungsbruch ein genomischer Prozess sein könnte.

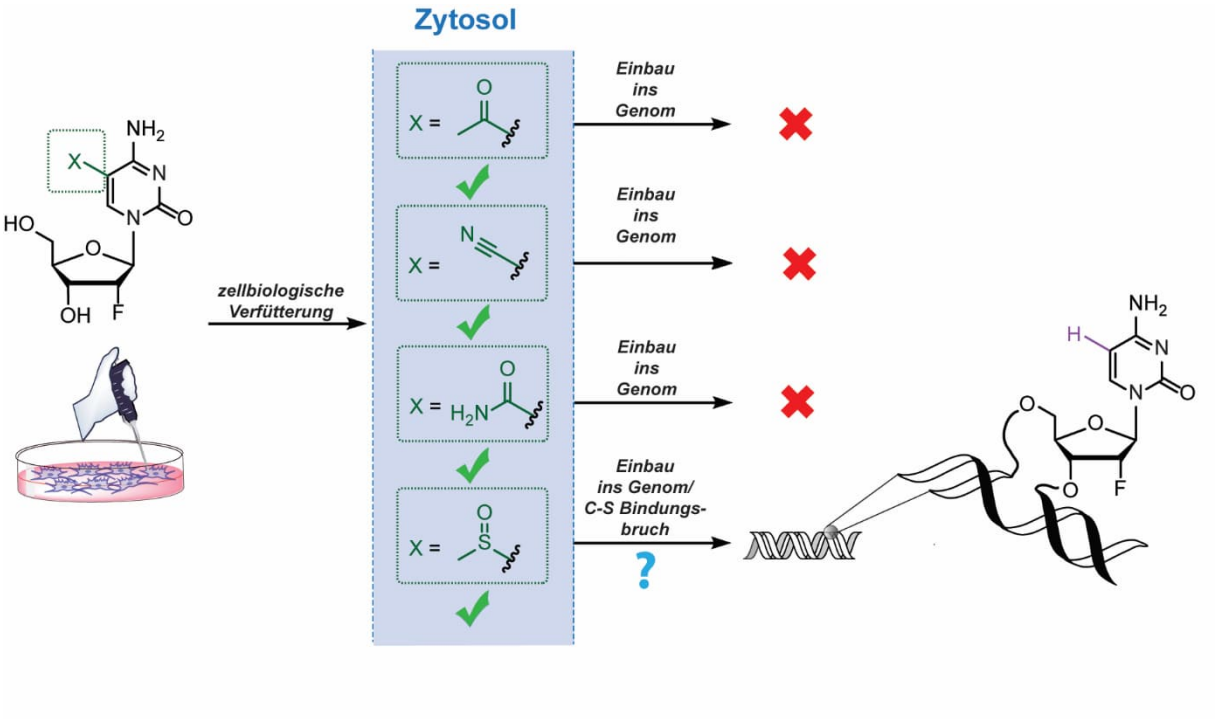
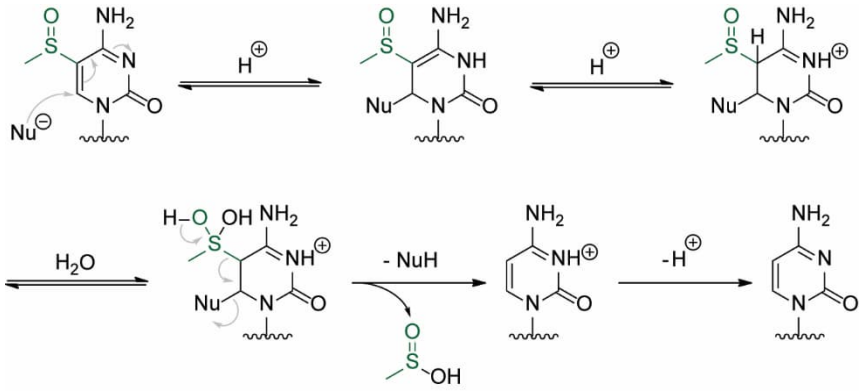


Abbildung 4.17: Ergebnisse der Verfütterungsstudien der vier verschiedenen Testmoleküle 2'-F-COMe-dC (**31**), 2'-F-CN-dC (**33**), 2'-F-$CONH_2$-dC (**45**) und 2'-F-SOMe-dC (**32**). Alle vier Nukleoside konnten im Zytosol nachgewiese werden. Nur im Fall von **32** wurde 2'-F-dC (**16**) im Genom detektiert.

Ein möglicher Desulfinylierungsmechanismus ist in **Schema 4.17** abgebildet. Für diesen Prozess wird analog zur Deformylierung ein nukleophiler Angriff an der C-6 Position der Nukleobase postuliert. Für die Deformylierungsreaktion wurde sowohl *in vitro* als auch *in vivo* gezeigt, dass der Angriff eines Nukleophils an der Nukleobase für den C-C-Bindungsbruch erforderlich ist.[93, 133, 199] Nach einem Michael-ähnlichen Angriff des Nukleophils an C-6 und anschließender Protonierung des *N*-3 erfolgt eine Imin-Enamin Tautomerie unter Protonierung von C-5. Nach einem nukleophilen Angriff von Wasser am elektrophilen Schwefel erfolgt der Austritt der Abgangsgruppe Methylsulfinsäure unter gleichzeitiger Abspaltung des Helfernukleophils und Rearomatisierung. Anschließende Deprotonierung führt zum desulfinylierten 2'-F-dC.



Schema 4.17: Potentieller Mechanimus einer Defunktionalisierung von 2'-F-SOMe-dC (**32**) durch einen C-S Bindungsbruch initiiert durch einen 1,4-artigen Angriff eines Nukleophils.

Eine Quantifizierung der aus der Desulfinylierung von 2'-F-SOMe-dC (**32**) resultierenden Produkte ergab für 2'-F-dC (**16**) eine Menge von $6.3x10^{-5}$/dN im Genom. Vergleichbare Mengen ($1.3x10^{-5}$/dN 2'-F-dC) wurden bei Verfütterung von 2'-F-fdC an N2a-Zellen detektiert. Dies lässt darauf schließen, dass eine potenzielle Desulfinylierung von 2'-F-SOMe-dC (**32**) mittels eines C-S-Bindungsbruchs eine ähnliche Effizienz wie die Deformylierung von 2'-F-fdC (**14**) aufweist. Fraglich ist jedoch, weshalb **32** selbst nicht in der genomischen DNA nachgewiesen werden konnte. Dies könnte darauf zurückzuführen sein, dass eine Desulfinylierungsreaktion im Vergleich zu Deformylierung schneller abläuft und das eingebaute 2'-F-SOMe-dC (**32**) zum untersuchten Zeitpunkt zum Großteil defunktionalisiert war und in nicht detektierbaren Mengen vorlag. Aus empirischen Erfahrungen kommt es bei Schwefel-haltigen Nukleosiden zu einer verstärkten Signalunterdrückung durch Ionensupression in MS Analysen vor. Analog zu 2'-F-cadC könnte das zu einer erschwerten Detektion des Analogons führen. Interessanterweise waren zum gemessenen Zeitpunkt von 72 h bereits signifikante Menge des remethylierten 2'-F-mdC detektiertbar. Die genomischen Level von 2'-F-mdC lagen bei Verfütterung von **32** bei $1.4x10^{-6}$/dN, was einem Anteil von 2.2% von 2'-F-dC entspricht. Im Vergleich dazu, wurden bei der Verfütterung von 2'-F-fdC an N2a zellen 2.5% der deformylierten 2'-F-fdC zu 2'-F-mdC remethyliert. Die Quantifizierungsdaten der im Genom detektierten 2'-fluorierten Nukeloside nach Verfütterung von 2'-F-SOMe-dC (**32**) sind in der **Abbildung 4.18** dargestellt.

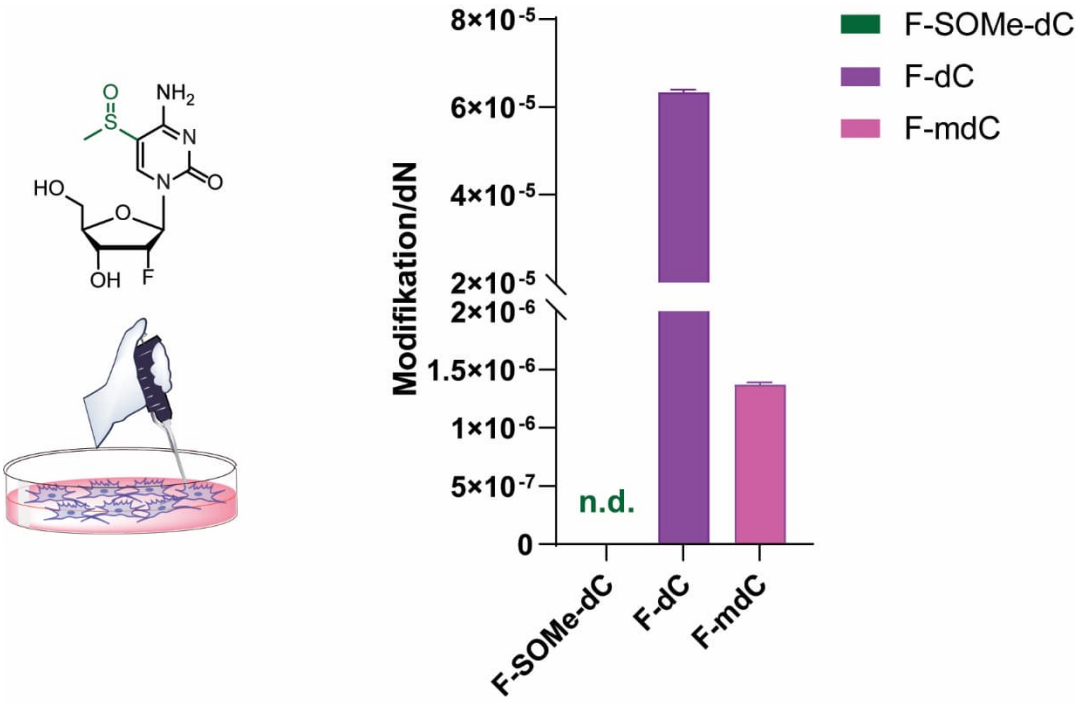


Abbildung 4.18: Darstellung der Quantifizierungsdaten der 2'-fluorierten Nukleoside im Genom der N2a Zellen nach Verfütterung von 2'-F-SOMe-dC (**32**). Die Daten repräsentieren ein technisches Triplikat. Es wurden absolute Mengen von 6.3×10^{-5}/dN 2'-F-dC/dN und 1.4×10^{-6}/dN 2'-F-mdC/dN im Genom detekiert. 2'-F-SOMe-dC konnte nicht detektiert werden (n.d.).

4.3.4 Zusammenfassung und Ausblick

In Anlehnung an die Struktur von 2'-F-fdC und 2'-F-cadC wurden im Rahmen dieser Arbeit vier verschiedene Nukleosid Analoga synthetisiert: 2'-F-COMe-dC (**31**), 2'-F-SOMe-dC (**32**), 2'-F-CN-dC (**33**) und 2'-F-$CONH_2$-dC (**45**). Bei diesen Nukleosid Derivaten handelt es sich um Moleküle welche im Vergleich zu 2'-F-fdC (**14**) und 2'-F-cadC (**15**) eine strukturell und elektronisch ähnliche Funktionalität an der C-5-Position der Nukleobase aufweisen. Die Synthese aller dieser neuartigen Nukleosidanaloga konnte ausgehend von 2'-fluoriertem 5-Iodo Desoxyuridin (**21**) in guten Ausbeuten durchgeführt werden. Die anschließenden zellbiologischen Verfütterungsexperimente an N2a Zellen zeigten, dass alle Testmoleküle in die Zelle aufgenommen wurden, da sie alle in der zytosolischen Fraktion detektiert wurden. Keines der Testmoleküle konnte in der genomischen DNA nachgewiesen werden. Im Falle von 2'-F-SOMe-dC (**32**) konnte jedoch desulfinyliertes 2'-F-dC sowie dessen Remethylierungsprodukt im Genom identifiziert und quantifiziert werden. Dieses Ergebnis legt die Vermutung nahe, dass beim 2'-F-SOMe-dC (**32**) ein intragenomischer C-S-Bindungsbruch stattfinden könnte. Eine weitere Untersuchung zur Verifikation des Einbaus von diesem 2'-F-fdC Analogon in die DNA könnte mit Hilfe von zeitaufgelösten Verfütterungsexperimenten weiter beleuchtet werden. Insgesamt legt die hier durchgeführte Studie nahe, dass eine Defunktionalisierung von 5-modifizierten dC Derivaten im Genom möglicherweise nicht nur auf

5-Formyl- und 5-Carboxyl-dC Deirvate beschränkt sein könnte. Die wesentliche Hürde für die Untersuchung des Prozesses scheint nicht die Defunktionalisierung der C-5 modifizierten dC Derivate zu sein, sondern die Detektion von deren Einbau ins Genom. Möglicherweise könnten auch andere funktionelle Gruppen als die natürlichen 5-Formyl- und 5-Carboxyl-dC Derivate Ziel dieses Defunktionalisierungsprozesses sein. Genauere Studien zum Sulfoxid Analogon des 2'-F-fdC hinsichtlich dessen Einbau und Defunktionalisierung im Genom insbesondere im Vergleich zum 2'-F-fdC könnten in diesem Zusammenhang durchaus von großem Interesse sein. Hierzu könnten zum Beispiel beide Testmoleküle in einem kompetitiven Verfütterungsexperiment eingesetzt werden, um zu zeigen, ob die Rate der Deformylierung durch die Answesenheit eines 2'-F-fdC Analogons beinträchtigt wird. Hierzu könnte eine Kombination von einem 2'-fluoriertem und einem isotopenmarkierten Testmolekül verwendet werden.

Im Bereich von 2'-F-cadC Analoga konnte ein Einbau von 2'-F-$CONH_2$-dC (**45**) in das Genom der N2a Zellen nicht nachgewiesen werden. Weitere Analoga, wie das von *Eveliina Ponkkonen* entwickelte 5-NO_2 Derivat des 2'-F-dC, konnten ebenfalls nicht in die genomische DNA der N2a Zellen eingebaut werden. Da aus der Studie mit 2'-F-cadC bekannt ist, dass nach einer Verfütterung des triphosphorylierten Nukleosids mit einem Nukleotid Transporter der Einbau ins Genom möglich und viel effizienter ist, könnte dieser Ansatz ebenfalls mit dem Nitro-substitiertem Nukleosid und den oben genannten Analoga in weiteren Experimeten untersucht werden. Insgesamt stellen die hier gezeigten fdC und cadC Analoga einen vielversprecheden Ansatzpunkt zur weiteren Untersuchung der Defunktionalisierung dieser epigenetisch bedeutenden DNA-Modifikationen dar.

4.3.5 Projektbeiträge

Im Rahmen dieser Arbeit habe ich die Entwicklung und Synthese der 2'-F-fdC und 2'-F-cadC Analoga 2'-F-COMe-dC (**31**), 2'-F-SOMe-dC (**32**), 2'-F-CN-dC (**33**) und 2'-F-$CONH_2$-dC (**45**) durchgeführt. Die Zellkulturversuche sowie die Isolation aller DNA-Proben und löslichem Nukleosidpool wurden von *Ewelina Kamińska* durchgeführt. Neben der Synthese der Testmoleküle führte ich die Entwicklung der massenspektrometrischen Methoden für die neuartigen Nukleoside, den enzymatischen Verdau, die Messung, Auswertung und Interpretation der Daten durch.

4.3.6 Material und Methoden

4.3.6.1 Synthesevorschriften und Analytik

Die Synthesevorschriften und Analytik sind im Kapitel 5.2.5 (2'-F-COMe-dC, **31**), Kapitel 5.2.6 (2'-F-SOMe-dC, **32**) , Kapitel 5.2.7 (2'-F-CN-dC, **33**) und Kapitel 5.2.8 (2'-F-$CONH_2$-dC, **45**) zu finden.

4.3.6.2 Enzymatischer Verdau

Der enzymatische Verdau der genomischen DNA und die Isolation des löslichen Nukleosidpool aus den Fütterungsexperimenten mit den hier beschriebenen Testmolekülen erfolgte wie im Kapitel 8.1 beschriebenen Protokoll.

4.3.6.3 UHPLC-MS/MS Analyse

Die quantitative UHPLC-MS/MS-Analyse der hydrolysierten DNA-Proben erfolgte wie im Kapitel 8.2 beschrieben mit einem 1290 UHPLC-System von Agilent, welches über einen UV-Detektor verfügt und mit einem 6490 Triple-Quadrupol Massenspektrometer von Agilent verbunden war. Die Chromatographie wurde mit einer Poroshell 120 SB-C8 Säule (Agilent, 2.7 µm, 2.1 mm × 150 mm) bei 35 °C mit einem Gradienten aus Wasser und Acetonitril mit jeweils 0.0075 % (v/v) Ameisensäure durchgeführt. Der Fluss betrug 0.35 mL/min: 0 →4 min; 0 →3.5% (v/v) MeCN; 4 →6.9 min; 3.5 →5 % MeCN; 6.9 →7.2 min; 5 →80 % MeCN; 7.2 →10.5 min; 80 % MeCN; 10.5 →11.3 min; 80 →0 % MeCN; 11.3 →13 min; 0 % MeCN. Die Quellparameter waren dieselben wie unter 8.2 beschrieben. Die für die jeweiligen Analyten spezifischen Parameter sind im Kapitel 5.3.3 in **Tabelle 5.3** bis **Tabelle 5.6** zusammengefasst. 2'-F-COMe-dC: **Tabelle 5.3.** 2'-F-SOMe-dC: **Tabelle 5.4**, 2'-F-CN-dC: **Tabelle 5.5**, 2'-F-$CONH_2$-dC: **Tabelle 5.6.**

4.3.6.4 Auswertung

Die Auswertung erfolgte wie im Kapitel 8.1 beschrieben. Die Kalibriergeraden der verschiedenen Modifikationen wurden aus bereits veröffentlichten Publikationen der Arbeitsgruppe verwendet.[133, 135, 188, 190] Die Kalibiriergeraden der quantifizierten Modifikationen sowie deren Gültigkeitsbereiche sind im im Kapitel 5.3.4 in **Tabelle 5.7** bis **Tabelle 5.9** zusammengefasst. Die Darstellung der Daten erfolgte mit Graphpad Prism und Adobe Illustrator.

5 EXPERIMENTELLER TEIL

5.1 Material und Methoden chemische Synthese

Chemische Synthese

Wenn nicht anders angegeben wurde, wurden alle Reaktionen unter Verwendung der Standard-Schlenktechnik durchgeführt, wobei Argon als Schutzgas verwendet wurde. Glasgeräte wurden vor jeder Reaktionsdurchführung mit einem Septum verschlossen, unter Hochvakuum mit einer Heizluftpistole bei 600 °C ausgeheizt und unter Vakuum abkühlen lassen. Feststoffe wurden unter einem leichten Argon Gegenstrom zugegeben.

Lösungsmittel

Für chemische Synthesen wurden über Molsieb getrocknete Lösemittel der Firma Acros verwendet, weche unter trockenen Bedingungen in einem Argon-Gegenstrom entnommen wurden. Für Extraktionen und Aufreinigungen mittels Säulenchromatographie wurden Lösemittel technischer Güte verwendet, welche vor Benutzung destilliert wurden.

Verwendete Chemikalien

Chemikalien wurden von den Firmen *Merck*, *Carbosynt*h, *Carbolution*, *TCI*, *Fluka*, *Acro*s und *ABCR* in Qualitätsstufen *puriss*, *p.a*. und *purum* bezogen und ohne weitere Aufreinigung verwendet.

Reaktionen mit einem Druckautoklav (CO-Insertion)

Für Synthese formylierter und carboxylierter Nukleoside wurde ein Druckautoklav mit einer CO-Atmosphäre von 3.5 bar verwendet. Bei dem Reaktor handelt es sich um ein *Miniclav*-Glas-Druckreaktor mit einem Borosilikas-Glas 3.3 der Firma *Büchi*. Der Reaktor wurde vor der Rektion im Ofen bei 100 °C ausgeheizt und unter einem leichten Argon-Strom abkühlen gelassen. Für die kontrollierte Zugabe von Lösungen während der Reaktion wurde eine Spritzenpumpe der Firma *KD Scientific* verwendet.

Dünnschichtchromatoraphie

Zur Kontrolle des Reaktionsvortschrittes, sowie zur Trennkontrolle bei säulenchromatographischen Aufreinigungen wurden DC-Platten F254 (Kieselgel auf

Aluminium) der Firma *Merck* verwendet. Aromatische, UV- und fluoreszenzaktive Substanzen wurden unter einer UV-Lampe (λ = 254 nm, 356 nm) beobachtet.

Für die Anfärbung der Substanzen auf der DC-Platte wurden folgende Lösungen verwendet:

- Ninhydrin (0.3 g), Eisessig (3.0 mL) und n-Butanol.
- $KMnO_4$ (3.0 g), K_2CO_3 (20 g) und KOH (0.30 g) in destilliertem H_2O (300 mL).
- Cer(IV)-ammoniummolybdat-Lösung, 120 g $(NH_4)_6Mo_7O_{24}\cdot 24H_2O$ und 5.0 g $Ce(NH_4)_4(NO_3)_6$ in 800 mL H_2SO_4, 10%-ig.

Normalphasen-Säulenchromatographie

Die säulenchromatographische Aufreinigung von Rohprodukten wurde in geeigneten Glassäulen unter variablen leichten Stickstoffüberdrücken nach der Flash-Technik durchgeführt. Als stationäre Phase wurde *Kieselgel 60* der Firma *Merck* (Korngröße: 0.063 nm–0.200 nm) verwendet. Angemessene Elutionsmittel wurden über Dünnschicht-chromatographie (DC) ermittelt. Fraktionen wurden in Regenzgläsern gesammelt und mit Hilfe von DC oder LC-MS analysiert.

Rotationsverdampfer

Zur Destillation der Lösemittel unter vermindetem Druck wurde ein Rotationsverdampfer *Laborota 4000* der Firma *Heidolph* verwendet. Sofern nicht abweichend angegeben, wurden Lösemittel bei 40 °C entfernt.

Bidestilliertes Wasser

Für HPLC-Aufreinigungen und für biochemische Experimente wurde Wasser aufgereinigt mit einer MiliQ-Anlage der Firma *Sartorius* verwendet.

***High Performance Liquid Chromatography* (HPLC)**

Zur Aufreinigung von Endverbindungen wurde ein semipräparatives HPLC-Trennverfahren an einem *Breeze-System* der Firma *Waters*, das aus einer *1525 Binary HPLC P*ump und einem *2487 UV/Vis Detektor* besteht, oder an einem *Agilent 1260 Infinity II 400 bar*-Pumpensystem mit einem *Agilent 1260 Infinity II VWD* bei 260 nm, durchgeführt. Als stationäre Phase wurden *VP 250/10 Nucleosil 100-7 C18* oder *Nucleodur* (250/10 mm, C18ec, 5 µm) *reversed-phase* Chromatographiesäulen der Firma *Macherey & Nagel* bei einer Flussrate von 5.0 mL/min verwendet. Für analytische Untersuchungen wurde ein *Waters Separationsmodul e2695* in Kombination mit einem 2487 UV/Vis Detektor (*Waters*) oder einem *Agilent 1260 Infinity II 800*

bar Flexible Prime-Pumpensystem mit *einem Agilent 1260 Infinity II MWD* bei 260 nm verwendet. Als stationäre Phase dienten *120/3 Nucleosil C18*-Chromatographiesäulen der Firma *Marcherey & Nagel* bei einer Flussrate von 0.5 mL/min oder 1.0 mL/min verwendet. Das für die HPLC Auftrennungen eingesetzten Laufmittelsystem bestanden aus doppelt destilliertem Wasser (ddH_2O) und Acetonitril (*HPLC-grade*). Falls gepufferte Laufmittelsysteme verwendet wurden, sind diese bei den jeweiligen Vorschriften genau angegeben.

pH-Meter

Ein pH-Meter *MP 220* der Firma *Mettler Toledo* wurde für pH-Messungen verwendet. Dabei wurde vor jeder Messung das Gerät mit einer pH = 4.0 und pH = 7.0 Standardlösung kalibriert.

Lyophilisation

Zum Entfernen der Lösemittel mittels Gefriertrockung wurden die Lyophille *2-4 LD plus* der Firma *Christ* verwendet. Die zu entfernenden Lösemittel wurden dabei zuvor mit flüssigem Stickstoff eingefroren.

Kernresonanzspektroskopie (NMR)

Eindimensionale ^{1}H-NMR-Spektren wurden größtenteils an einem *Bruker Ascend 400 NMR*-Spektrometer aufgenommen. Eindimensionale ^{1}H-, ^{13}C-, ^{19}F- sowie ^{15}N-NMR-Spektren sowie zweidimensionale ^{1}H- und ^{13}C-Spektren (COSY, HSQC und HMBC) zur genauen Charakterisierung der hergestellten Verbindungen wurden von der Abteilung für NMR-Spektroskopie des Hauses F des Departments Chemie und Pharmazie der LMU München, namentlich von *Dr. David Stephenson* und *Claudia Dubler*, durchgeführt. Bei diesen Messungen wurden als Spektrometer die Geräte Bruker Avance III HD 400, *Varian NMR-System 600* und *Bruker Avance III HD* mit *Cryo-Kopf 800* verwendet. Die chemischen Verschiebungen sind in ppm (*parts per million*) angegeben, welche sich auf die Mitte des jeweiligen Kopplungsmusters beziehungsweise auf die Ausdehnung des Multipletts beziehen. Als interner Standard relativ zu den chemischen Verschiebungen dienten die verwendeten deuterierten Lösemittel. Die Multiplizitäten der Signale werden als s (Singulet), d (Duplet), t (Triplet), q (Quartet), m (Multiplet) bzw. deren Kombinationen angegeben. Außerdem ist zu beachten, dass die Angaben der Multiplizitäten phänomenologisch sind. Das heißt, dass die Daten entsprechend ihrer Erscheinung und nicht ihrer theoretisch zu erwartenden Multiplizität angegeben wurden. Kopplungskonstanten sind in Hertz (Hz) angegeben und beziehen sich, sofern nicht anders angegeben, auf H-H-Nah- und Fernkopplungen. Im Fall, dass Signale nicht

eindeutig zugeornet werden konnten und austauschbar sind, wurden die zugehörigen Atome mit “*” / “**” oder “ǀ” / “ǁ” gekennzeichnet. NMR Spektren wurden mit der Software *Mestrenova* ausgewertet. Die angegebene Nummerierung der Atome folgt nicht entsprechenden IUPAC Regeln.

Massenspektrometrie (LC-MS, HR-ESI, HR-EI, Orbitrap, QQQ-MS, MALDI)

Niederaufgelöste ESI-Massenspektrometrie zur Reaktionskontrolle wurde über Direkteinspritzung bzw. mit Flüssigchromatographie-Kopplung (LC-MS) an einem Massenspektrometer *MSQ Plus* der Firma *Thermo Scientific* durchgeführt. Dem genannten Massenspektrometer ist ein Flüssigchromatographie- sowie Einspritzsystem *Diones Ulimate3000* des gleichen Herstellers vorgeschaltet. Hochaufgelöste Massenspektrometrie (HR-ESI/HR-EI-MS) wurde von der Abteilung für Massenspektrometrie des Departments für Chemie udn Pharmazie der LMU München, namentlich von *Dr. W. Spahl* und *S. Kosak* durchgeführt. Die Spektren wurden mit einem *LTQ-FTI-CR* der Firma *Finnigan* aufgenommen. EI Spektren wurden mit einer *QExactive* der Firma *Thermo Fisher* angefertigt.

Für die Analyse von Oligonukleotiden per MALDI-TOF-Massenspektrometrie wurde ein *Bruker Autoflex II* verwendet. Die Proben wurden vor der Messung unter Verwendung von 0.025 um Filtern (*Millipore*) entsalzen und anschließend mit einer HPA-Matrix (3-Hydroxypicolinsäure) cokristallisiert.

Infrarotspektroskopie (IR)

Aufnahme von IR-Spektren erfolgte mit einem *BX FT-IR* System von *Perkin-Elmer* mit einer Diamant-ATR (*attenuated total reflection*) Einheit. Feststoffe wurden dabei entweder zuvor in wenig Dichlormethan gelöst, ein Tropfen der Lösung auf die ATR-Einheit gegeben und das Lösemittel verdampft, oder wurden die Feststoffe direkt auf die ATR-Einheit gegeben und mit einem Stempel auf der ATR-Einheit verdichtet Öle und flüssige Verbidnungen wurden als Filme aufgetragen. Die Wellenzahlen sind in cm^{-1} angegeben und die Intensitäten der erhaltenen Signale mit vs (sehr stark), s (stark), m (mittel), w (schwach) oder b (breit) beschrieben.

UV/Vis-Spektroskopie

Zur Aufnahme von UV/Vis-Spetren wurde ein *JASCO V650* Spektrometer verwendet. Es wurden 100 Scans/min bei einem Wellenlängenintervall von 0.5 nm verwendet. Das Lösemittel, in dem die zu analysierende Verbidnung gelöst war, wurde vor der Messung als Blindprobe (*blank*) verwendet.

5.2 Synthesevorschriften und Analytik

5.2.1 Allgemeine Reakionsvorschriften

5.2.1.1 Allgemeine Reakionsvorschrift I

In einem Polypropylen Rohr (*Falcon tube*) wurde eine TBS-geschützte Verbindung (1.0 Äq.) in EtOAc (15 mL/mmol) gelöst und mit HF • Pyridin (70%, 15.0 Äq.) versetzt. Die Reaktionsmischung wurde 18 h bei Raumtemperatur gerührt, anschließend auf 0 °C gekühlt, mit TMS • OMe (32.0 Äq.) versetzt und für 1 h bei 0 °C gerührt. Flüchtigen Bestandteile wurden in vacuo entfernt, das erhaltene Rohprodukt in H_2O aufgenommen und mit CH_2Cl_2 extrahiert. Die wässrige Phase wurde unter vermindertem Druck konzentriert und mittels semipräparativer HPL-Chromatographie aufgereinigt um ein freies Nukleosid zu erhalten.

5.2.1.2 Allgemeine Reaktionsvorschrift II

Das Reaktionsgefäß eines Hochdruck-CO-Autoklaven wurde mit einem iodierten Nukleosid (1.0 Äq.) und $[Pd(MeCN)_2Cl_2]$ (0.05 Äq.) beladen und im geeigneten Alkohol ROH gelöst (bei flüssigen Alkoholen) (20 mL/mmol) oder mit einer Lösung vom geeigneten Alkohol (5.0 Äq.) ROH in Toluol (15 mL/mmol) versetzt. Der Reaktor wurde unter Rühren mit CO (3.5 bar) gespült um zu testen, ob die Apparatur dicht ist. Anschließend wurde der Reaktor zwei Mal mit CO (3.5 bar) gespült. Anschließend wurde DIPEA (2.0 Äq.) wurde zugegeben. Der Autoklav wurde erneut unter eine CO-Druckatmosphäre (3.5 bar) versetzt und die Reaktionsmischung wurde 18 h bei 65 °C gerührt. Das erhaltene Reaktionsgemisch wurde mithilfe einer Fritte über SiO_2 filtriert und mit Ethylacetat (100 mL) gewaschen. Nachdem die Lösungsmittel *in vacuo* entfernt wurden, wurde das Rohprodukt säulenchromatographisch auf gereinigt

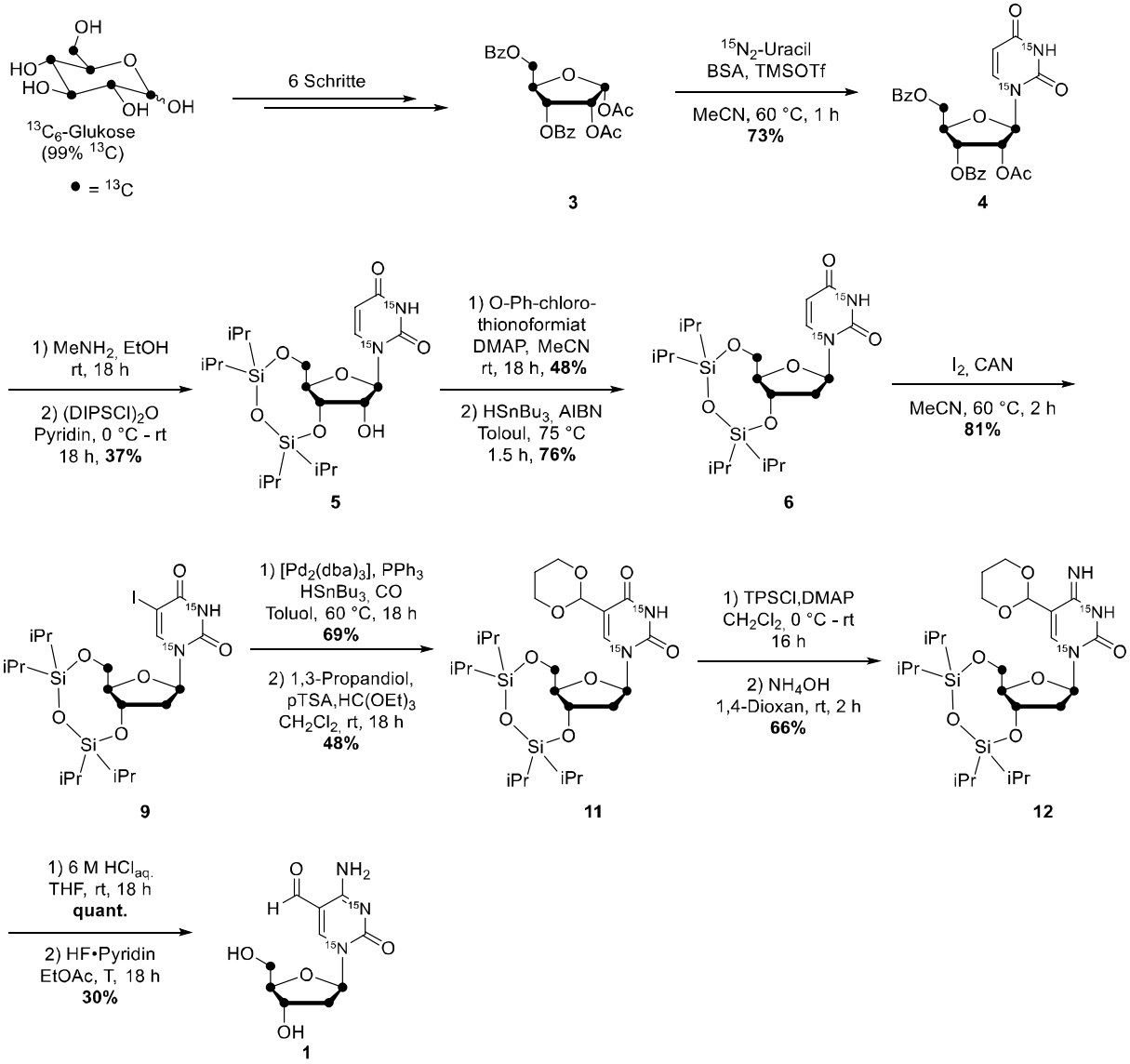
5.2.1.3 Allgemeine Reaktionsvorschrift III

Nach einer modifzierten Literaturvorschrift[93] wurde 1,2,4-Triazol (9.3 Äq.) in MeCN (11 mL) gelöst, auf 0 °C gekühlt und 10 min bei 0 °C gerührt. $POCl_3$ (2.3 Äq.) wurde langsam zugetropft und die Reaktionsmischung wurde weitere 10 min bei 0 °C gerührt. Anschließend wurde NEt_3 (8.7 Äq.) tropfenweise hinzugegeben und die Mischung wurde 20 min bei 0 °C gerührt. Im nächsten Schritt wurde eine Lösung von der zu aminierenden Verbindung (1.0 Äq.) in MeCN (5 mL/mmol) langsam hinzugetropft. Die Reaktionsmischung wurde bei Raumtemperatur gerührt und anschließend durch Zugabe einer gesättigten wässrigen $NaHCO_3$-Lösung

neutralisiert. MeCN wurde unter vermindertem Druck entfernt und die Lösung mit CH_2Cl_2 extrahiert. Die vereinigten organischen Phasen wurden über Na_2SO_4 getrocknet und die Lösungsmittel wurden *in vacuo* entfernt.

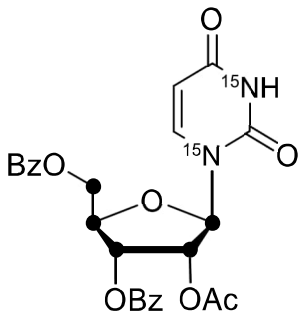
Das erhaltene Triazolderivat wurde in 1,4-Dioxan (20 mL/mmol) gelöst und tropfenweise mit einer wässrigen Lösung von NH_3 (25%, 5 mL/mmol) versetzt. Die Reaktionsmischung wurde bei Raumtemperatur gerührt, durch Zugabe einer gesättigten wässrigen NH_4Cl-Lösung beendet und mit EtOAc extrahiert. Die vereinigten organischen Phasen wurden mit einer gesättigten wässrigen NaCl-Lösung gewaschen und anschließend über Na_2SO_4 getrocknet. Lösungsmittel wurden *in vacuo* entfernt und das Rohprodukt wurde säulenchromatographisch aufgereinigt um das aminierte Nukleosid zu erhalten.

5.2.2 Synthese von 5-Formyl-2'-desoxy-(1',2',3',4',5'-$^{13}C_5$,N^1,N^3-$^{15}N_2$)-cytidin (1)



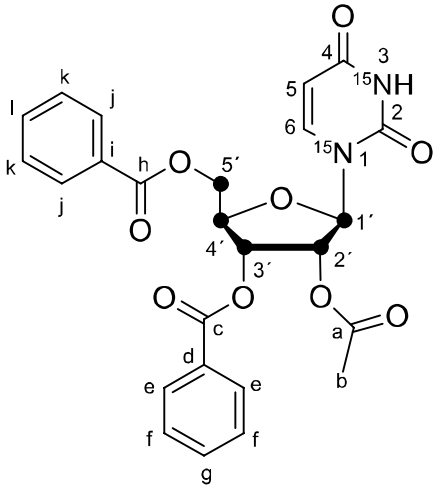
Schema 5.1: Syntheseschema des $^{13}C_5$-$^{15}N_2$-markierten fdC (**1**) ausgehend von komerzielle erwerblichemn $^{13}C_6$-Glukose.

5.2.2.1 2'-*O*-Acetyl-3',5'-di-*O*-benzoyl-(1',2',3',4',5'-$^{13}C_5$,*N*^{1},*N*^{3}-$^{15}N_2$)-uridin (4)



4

Die für diese Synthese verwendete $^{13}C_5$-1,2-Bis-*O*-acetyl-3,5-bis-*O*-benzoyl-β-D-ribofuranose (**3**) wurde in sechs literaturbekannten Schritten aus $^{13}C_6$-Glukose hergestellt.[185] Die isotopenmarkierte Ribose (**3**, 0.50 g, 1.12 mmol, 1.0 Äq.) und $^{15}N_2$-Uracil (0.13 g, 1.29 mmol, 1.01 Äq.) wurden in Acetonitril (6.3 mL) gelöst, mit *N*,*O*-Bis-TMS-acetamid (0.68 g, 0.82 mL, 3.35 mmol, 3.0 Äq.) versetzt, 15 min bei Raumtemperatur und 15 min bei 60 °C gerührt. Anschließend wurde TMSOTf (0.87 g, 0.71 mL, 3.91 mmol, 3.5 Äq) zugetropft und die Reaktionslösung wurde 30 min bei 60 °C gerührt. Nach Abkühlen auf Raumtemperatur wurde die Reaktion durch Zugabe einer gesättigten wässrigen $NaHCO_3$ (30 mL) beendet. Die wässrige Phase wurde mit CH_2Cl_2 (5 x 100 mL) extrahiert. Die vereinigten organischen Phasen wurden mit einer gesättigten NaCl-Lösung (100 mL) gewaschen und über Na_2SO_4 getrocknet. Alle flüchtigen Bestandteile wurden unter vermintertem Druck entfernt und das Rohprodukt wurde säulenchromatographisch (SiO_2, *i*Hexan/EtOAc = 1:1 → *i*Hexan/EtOAc = 1:2) aufgereinigt um das Nukleoisd **4** (0.40 g, 0.80 mmol, 73%) in Form eines gelben Öls zu erhalten.



.**R_f** = 0.40 (SiO_2, *i*Hexan/EtOAc = 1:1).

^{1}H NMR (600 MHz, $CDCl_3$): δ/ppm = 8.21 (d, $^1J_{N-H}$ = 91.4 Hz, 1H, N3-H), 8.08-8.05 (m, 4H, C_{e-f}-H)*, 7.65-7.60 (m, 2H, C_g-H, C_l-H), 7.51-7.47 (m, 4H, C_{j-k}-H)*, 7.34 (dt, ^3J = 8.2 Hz, $^3J_{N-H}$ = 2.4 Hz, 1H, C5-H), 6.20 (d, $^1J_{C-H}$ = 168.7 Hz, 1H, C1'-H), 5.74 (d, $^1J_{C-H}$ = 159.1 Hz, 1H, C3'-H), 5.58-5.57 (m, 1H, C6-H), 5.54 (d, $^1J_{C-H}$ = 154.6 Hz, 1H, C2'-H), 4.79 (d, $^1J_{C-H}$ = 151.4 Hz, 1H, C4'-H), 4.79 (d, $^1J_{C-H}$ = 151.4 Hz, 1H, C4'-H), 4.63 (d, $^1J_{C-H}$ = 151.7 Hz, 1H, C5'-H), 2.06 (s, 3H, C_b-H_3).

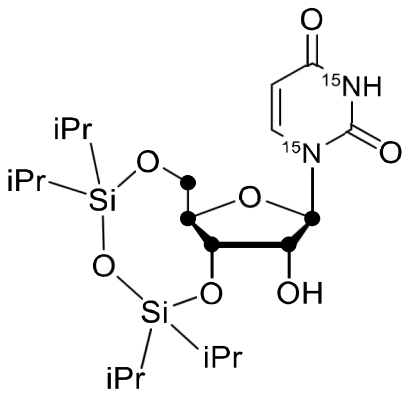
^{13}C-NMR (126 MHz, $CDCl_3$): δ/ppm = 169.8 (C_a), 166.1 (C_c)*, 165.5 (C_h)*, 162.2 (d, $^1J_{C-N}$ = 9.3 Hz, C4), 150.0 (t, $^1J_{C-N}$ =18.2 Hz, C2), 139.3 (d, $^2J_{C-N}$ = 12.0 Hz, C5), 134.1 (2C,C_e)**, 133.9 (2C,C_f)**, 130.0 (2C,C_j)**, 129.8 (2C,C_k)**, 128.9 (2C,C_g ,C_l)**, 128.8 (2C, C_d ,C_i)**, 105.7 (d, $^1J_{C-N}$ = 33.4 Hz, C-6) 87.4 (dd, *J* = 45.3 Hz, *J* =13.7 Hz, C1'), 80.4 (t, *J* = 40.3 Hz, C4'), 73.1 (t, *J* = 42.0 Hz, C2'), 70.9 (t, *J* = 39.5 Hz, C3'), 63.8 (d, *J* = 42.0 Hz, C5'), 20.6 (C_b).

^{15}N-NMR (41 MHz, $CDCl_3$): *δ* (ppm) = -224.9, -255.0.

HRMS (ESI): berechnet für $C_{20}{}^{13}C_5H_{21}{}^{15}N_2O_9^-$ [(M-H)$^-$]: 500.1361, gefunden 500.1364.

Die gemessenen Daten stimmen weitgehend mit der Literatur überein.[135]

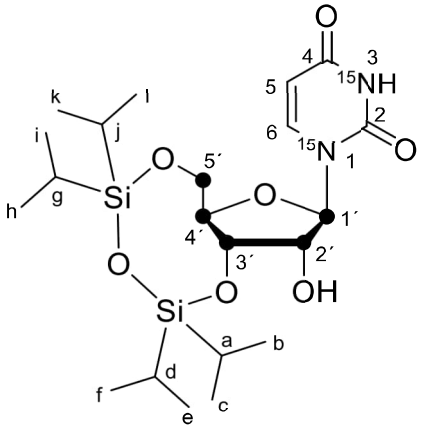
5.2.2.2 3',5'-(1,1,3,3,-Tetraisopropyldisiloxan-1,3-diyl)-(1',2',3',4',5'-$^{13}C_5$,*N*^{1},*N*^{3}-$^{15}N_2$)-uridin (5)



5

Nukleosid **4** (0.43 mg, 0.86 mmol, 1.0 Äq.) wurde in einer ethanolischen $MeNH_2$-Lösung (33%, 5.4 mL) gelöst und 20 h bei Raumtemperatur gerührt. Die Reaktionslösung wurde unter vermindertem Druck einkonzentriert, in H_2O (30 mL) aufgenommen und so oft mit CH_2Cl_2 extrahiert bis die organischen Extrakte kein UV Signal auf der DC Platte gezeigt haben (7 x 50 mL). Die wässrige Lösung wurde lyophilisiert und das in Form eines farblosen Pulvers erhaltene freie **4a** Nukleosid wurde ohne Aufreinigungsschritte für den nächsten Synthesschritt eingesetzt.

Verbindung **4a** (0.19 mg, 0.75 mmol, 1.0 Äq.) wurde in Pyridin (5 mL) gelöst, auf 0 °C abgekühlt und tropfenweise mit 1,3-Dichloro-1,1,3,3-tetraisopropyldisiloxan (0.26 g, 0.26 mL, 0.82 mmol, 1.1 Äq) versetzt und 18 h bei Raumtemperatur gerührt. Pyridin wurde *in vacuo* entfernt, der Rückstand wurde in CH_2Cl_2 (50 mL) aufgenommen und mit einer gesättigten wässrigen NH_4Cl (20 mL) versetzt. Die wässrige Phase wurde mit CH_2Cl_2 (4 x 50 mL) extrahiert. Die vereinigten organischen Phasen wurden mit einer gesättigten NaCl-Lösung (50 mL) gewaschen und über Na_2SO_4 getrocknet. Alle flüchtigen Bestandteile wurden unter vermintertem Druck entfernt und das Rohprodukt wurde säulenchromatographisch (SiO_2, *i*Hexan/EtOAc = 2:1 → *i*Hexan/EtOAc = 1:1) aufgereinigt um das Nukleoisd **5** (0.16 g, 0.33 mmol, 37%) in Form eines gelben Öls zu erhalten.



R_f = 0.29 (SiO_2, *i*Hexan/EtOAc = 1:1).

^{1}H NMR (600 MHz, $CDCl_3$): δ/ppm = 8.05 (d, $^1J_{N\text{-}H}$ = 90.3 Hz, 1H, N3-H), 7.66 (dt, ^3J = 8.1 Hz, $^3J_{N\text{-}H}$ = 2.4 Hz, 1H, C5-H), 5.72 (d, $^1J_{C\text{-}H}$ = 173.3 Hz, 1H, C1'-H), 5.70-5.67 (m, 1H, C6-H), 4.39 (d, $^1J_{C\text{-}H}$ = 150.8 Hz, 1H, C2'-H), 4.27-3.91 (m, 4H, C3'-H, C4'-H, C5'-H), 1.11-0.98 (m, 28H, $C_{a\text{-}l}$-H).

^{13}C-NMR (126 MHz, $CDCl_3$): δ/ppm = 162.7 (d, $^1J_{C\text{-}N}$ = 9.8 Hz, C4), 149.7 (t, $^1J_{C\text{-}N}$ =19.2 Hz, C2), 140.2 (d, $^2J_{C\text{-}N}$ = 13.7 Hz, C5), 103.7 (C6). 91.0 (dd, *J* = 43.0 Hz, *J* =13.4 Hz, C1'), 82.2 (t,

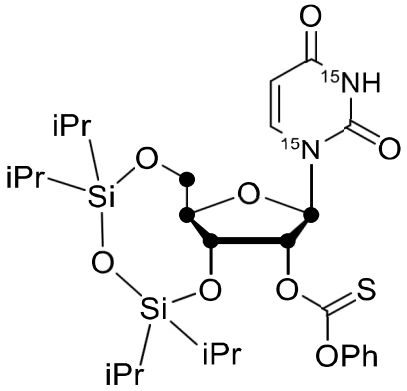
J = 42.1 Hz, C2'), 75.3 (t, *J* = 40.6 Hz, C3'), 69.3 (t, *J* = 39.9 Hz, C4'), 60.5 (d, *J* = 43.7 Hz, C5'), 17.6 (C_b)*, 17.5 (C_c)*, 17.4 (C_e)*, 17.4 (C_f)*, 17.2 (C_h)*, 17.1 (C_i)*, 17.0 (C_k)*, 17.0 (C_l)*, 13.5 (C_a)**, 13.0 (C_d)**, 13.0 (C_j)**, 12.7 (C_g)**.

^{15}N-NMR (41 MHz, $CDCl_3$): δ (ppm) = -224.4, -224.3.

HRMS (ESI): berechnet für $C_{16}{}^{13}C_5H_{39}{}^{15}N_2O_7Si_2^+$ [$(M+H)^+$]: 494.2399, gefunden 494.2406.

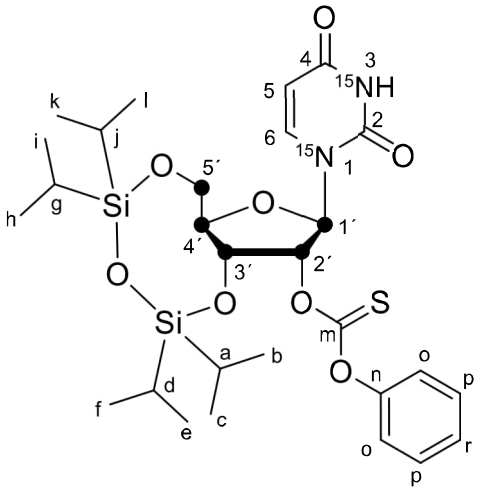
Die gemessenen Daten stimmen weitgehend mit der Literatur überein.[135]

5.2.2.3 3',5'-(1,1,3,3,-Tetraisopropyldisiloxan-1,3-diyl)-2'-*O*-phenoxy-thiocarbonyl (1',2',3',4',5'-$^{13}C_5$,*N*^{1},*N*^{3}-$^{15}N_2$)-uridin (46)



46

Verbindung **5** (0.16 mg, 0.32 mmol, 1.0 Äq.) wurde in MeCN (4.5 mL) gelöst, mit DMAP (0.11 g, 0.94 mmol, 2.9 Äq) und *O*-Phenyl-Chlorothionoformiat (80.0 mg, 70.0 µmol, 1.5 Äq) versetzt und 30 h bei Raumtemperatur gerührt. Die Reaktionslösung wurde mit einer gesättigten wässrigen NH_4Cl (10 mL) versetzt, MeCN wurde *in vacuo* entfernt und die wässrige Lösung wurde mit CH_2Cl_2 (5 x 20 mL) extrahiert. Die vereinigten organischen Phasen wurden mit einer gesättigten NaCl-Lösung (50 mL) gewaschen und über Na_2SO_4 getrocknet. Alle flüchtigen Bestandteile wurden unter vermintertem Druck entfernt und das Rohprodukt wurde säulenchromatographisch (SiO_2, *i*Hexan/EtOAc = 4:1 → *i*Hexan/EtOAc = 2:1) aufgereinigt um das Nukleoisd **46** (97.0 mg, 0.15 mmol, 48%) in Form eines farblosen Schaums zu erhalten.



R_f = 0.29 (SiO_2, *i*Hexan/EtOAc = 2:1).

^{1}H NMR (600 MHz, $CDCl_3$): δ/ppm = 8.05 (d, $^1J_{N\text{-}H}$ = 92.3 Hz, 1H, N3-H), 7.73 (dt, ^3J = 8.0 Hz, $^3J_{N\text{-}H}$ = 2.2 Hz, 1H, C5-H), 7.42 (d, ^3J = 8.2 Hz, 2H, C_o-H)*, 7.31 (t, ^3J = 7.16 Hz, 2H, C_r-H), 7.12 (^3J = 8.2 Hz, 2H, C_p-H)* 6.01 (td, $^1J_{C\text{-}H}$ = 162.9 Hz, ^3J = 5.1 Hz 1H, C2'-H), 6.15-5.69 (m, 2H, C6-H, C1'-H), 4.74-3.45 (m, 4H, C3'-H, C4'-H, C5'-H), 1.11-1.01 (m, 28H, $C_{a\text{-}l}$-H).

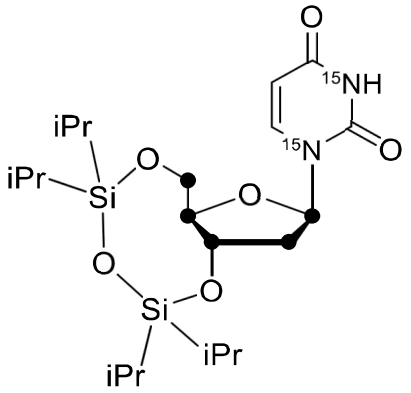
^{13}C-NMR (126 MHz, $CDCl_3$): δ/ppm = 193.7 (C_m), 162.5 (d, $^1J_{C\text{-}N}$ = 9.58 Hz, C4), 153.4 (C_n), 149.3 (t, $^1J_{C\text{-}N}$ =18.5 Hz, C2), 139.4 (d, $^2J_{C\text{-}N}$ = 12.8 Hz, C5), 129.6 (2C, C_o)*, 126.7 (C_r), 121.7 (2C, C_p)*, 102.3 (C6), 88.6 (dd, *J* = 43.0 Hz, *J* = 53.5 Hz, $J_{C\text{-}N}$ = 12.2 Hz, C1'), 83.7 (t, *J* = 41.9 Hz, C2'), 82.2 (t, *J* = 42.3 Hz, C3'), 68.1 (t, *J* = 41.0 Hz, C4'), 59.4 (d, *J* = 43.5 Hz, C5'), 17.4 (C_b)*, 17.4 (C_c)*, 17.3 (C_e)*, 17.3 (C_f)*, 17.2 (C_h)*, 17.2 (C_i)*, 17.0 (C_k)*, 17.0 (C_l)*, 13.4 (C_a)**, 13.0 (C_d)**, 13.0 (C_j)**, 12.4 (C_g)**.

^{15}N-NMR (41 MHz, $CDCl_3$): *δ* (ppm) = -223.5, -223.5.

HRMS (ESI): berechnet für $C_{23}{}^{13}C_5H_{43}{}^{15}N_2O_8SSi_2^+$ [(M+H)$^+$]: 630.2382, gefunden 630.2396.

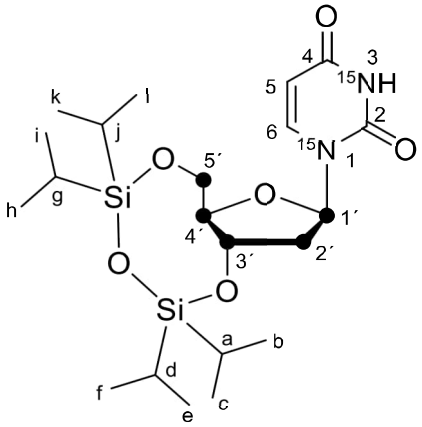
Die gemessenen Daten stimmen weitgehend mit der Literatur überein.[135]

5.2.2.4 3',5'-(1,1,3,3,-Tetraisopropyldisiloxan-1,3-diyl)-2'-desoxy-(1',2',3',4',5'-$^{13}C_5$,*N*^{1},*N*^{3}-$^{15}N_2$)-uridin (6)



6

Nukleosid **46** (97.0 mg, 0.15 mmol, 1.0 Äq.) wurde in Toluol (3 mL) gelöst, mit AIBN (5.0 mg, 30.8 µmol, 0.2 Äq) und $HSnBu_3$ (67.0 mg, 62.0 µmol, 1.5 Äq) versetzt und nach der *Freeze-Pump-Thaw* Technik entgast. Anschließend wurde die Reaktionslösung 1.5 h auf 75 °C erwärmt, auf Raumtemperatur abkühlen lassen und mit H_2O (15 mL) versetzt. Die wässrige Phase wurde mit CH_2Cl_2 (4 x 30 mL) extrahiert. Die vereinigten organischen Phasen wurden mit einer gesättigten NaCl-Lösung (50 mL) gewaschen und über Na_2SO_4 getrocknet. Alle flüchtigen Bestandteile wurden unter vermintertem Druck entfernt und das Rohprodukt wurde säulenchromatographisch (SiO_2, *i*Hexan/EtOAc = 2:1) aufgereinigt um das Nukleoisd **6** (56.0 mg, 0.12 mmol, 76%) in Form eines farblosen Öls zu erhalten.



R_f = 0.34 (SiO_2, *i*Hexan/EtOAc = 2:1).

^{1}H NMR (600 MHz, $CDCl_3$): δ/ppm = 9.60 (d, $^1J_{N\text{-}H}$ = 90.4 Hz, 1H, N3-H), 7.78 (dt, ^3J = 8.1 Hz, $^3J_{N\text{-}H}$ = 2.2 Hz, 1H, C5-H), 6.04 (d, $^1J_{C\text{-}H}$ = 174.0 Hz, ^3J = 5.6 Hz 1H, C1'-H), 5.70-5.68 (m, 1H,

C6-H), 4.51-3.55 (m, 5H, C3'-H, C4'-H, C5'-H), 3.82 (d, $^1J_{C\text{-}H}$ = 133.3 Hz, 1H, C2'-H_a)*, 2.26 (d, $^1J_{C\text{-}H}$ = 133.3 Hz, 1H, C2'-H_b)*, 1.08-0.99 (m, 28H, $C_{a\text{-}l}$-H).

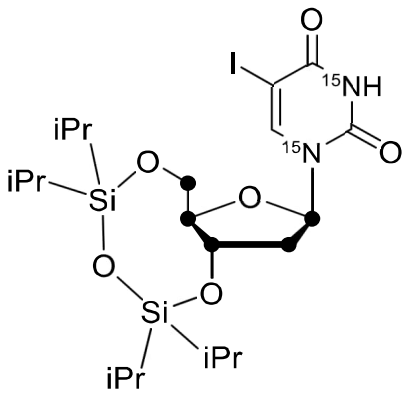
^{13}C-NMR (126 MHz, $CDCl_3$): δ/ppm = 162.9 (d, $^1J_{C\text{-}N}$ = 9.4 Hz, C4), 149.8 (t, $^1J_{C\text{-}N}$ =18.5 Hz, C2), 139.6 (d, $^2J_{C\text{-}N}$ = 13.0 Hz, C5), 101.6 (C6). 85.1 (t, *J* = 42.0 Hz, C4'), 84.3 (dd, *J* = 36.4 Hz, $^1J_{C\text{-}N}$ = 10.4 Hz C1'), 67.0 (t, *J* = 39.1 Hz, C3'), 60.0 (d, *J* = 37.0 Hz, C5'), 39.9 (t, *J* = 37.0 Hz, C2'), 17.5-16.7 (8C, $C_{b\text{-}i}$), 13.4 (C_a)**, 12.9 (C_d)**, 12.8 (C_j)**, 12.4 (C_g)**.

^{15}N-NMR (41 MHz, $CDCl_3$): *δ* (ppm) = -224.1, -224.2.

HRMS (ESI): berechnet für $C_{16}{}^{13}C_5H_{39}{}^{15}N_2O_6Si_2^+$ [$(M+H)^+$]: 478.2450, gefunden 478.2451.

Die gemessenen Daten stimmen weitgehend mit der Literatur überein.[135]

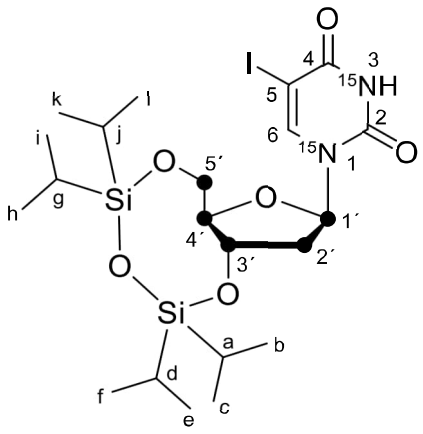
5.2.2.5 3',5'-(1,1,3,3,-Tetraisopropyldisiloxan-1,3-diyl)-5-iodo-2'-desoxy-(1',2',3',4',5'-$^{13}C_5$,*N*^{1},*N*^{3}-$^{15}N_2$)-uridin (9)



9

Die Reaktion wurde nach einer modifzierten Literaturvorschrift durchgeführt.[200] Nukleosid **6** (0.45 g, 0.94 mmol, 1.0 Äq.) wurde in Acetonitril (56 mL) gelöst, mit Iod (0.47 g, 1.88 mmol, 2.2 Äq.) und Ammoniumcer(IV)-nitrat (1.03 g, 1.88 mmol, 2.2 Äq.) versetzt. Die Reaktionslösung wurde für 2.5 h bei 60 °C gerührt. Anschließend wurde die Reaktion durch Zugabe einer gesättigten wässrigen $NaHCO_3$-Lösung (50 mL) und einer gesättigten wässrigen $Na_2S_2O_3$-Lösung (50 mL) beendet. Acetonitril wurde *in vacuo* entfernt und die wässrige Phase wurde mit CH_2Cl_2 (5 x 50 mL) extrahiert. Die vereinigten organischen Phasen wurden mit einer gesättigten NaCl-Lösung (150 mL) gewaschen und über Na_2SO_4 getrocknet. Nach Entfernen

aller flüchtigen Bestandteile unter vermintertem Druck wurde das Rohprodukt säulenchromatographisch (SiO_2, *i*Hexan/EtOAc = 4:1) aufgereinigt um das Nukleoisd **9** (0.46 g, 0.76 mmol, 81%) in Form eines farblosen Schaums zu erhalten.



R_f = 0.60 (SiO_2, *i*Hexan/EtOAc = 2:1).

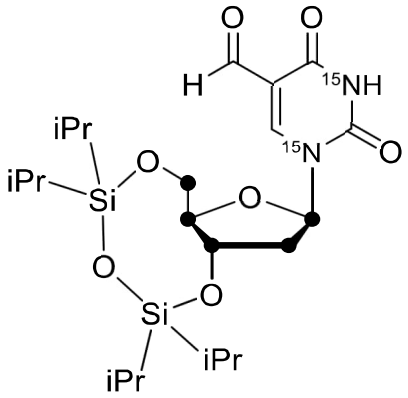
^{1}H NMR (600 MHz, $CDCl_3$): δ/ppm = 8.45 (dd, $^1J_{N\text{-}H}$ = 92.2 Hz, $^3J_{N\text{-}H}$ = 2.2 Hz 1H, N3-H), 8.02 (t, $^3J_{N\text{-}H}$ = 2.0 Hz, 1H, C6-H), 5.99 (d, $^1J_{C\text{-}H}$ = 174.1 Hz, ^3J = 6.4 Hz 1H, C1'-H), 4.47 (d, $^1J_{C\text{-}H}$ = 133.0 Hz 1H, C3'-H), 4.13 (d, $^1J_{C\text{-}H}$ = 140.0 Hz 1H, C4'-H), 4.01 (dd, $^1J_{C\text{-}H}$ = 140.0 Hz, ^2J = 13.6 Hz 1H, C5'-H_a)*, 3.77 (d, $^1J_{C\text{-}H}$ = 146.0 Hz ,1H, C5'-H_b)*, 2.50 (d, $^1J_{C\text{-}H}$ = 133.2 Hz, 1H, C2'-H_a)**, 2.28 (d, $^1J_{C\text{-}H}$ = 139.5 Hz, 1H, C2'-H_b)**, 1.12-1.01 (m, 28H, $C_{a\text{-}i}$-H).

^{13}C-NMR (126 MHz, $CDCl_3$): δ/ppm = 175.0 (d, $^1J_{C\text{-}N}$ = 9.4 Hz, C4), 150.1 (t, $^1J_{C\text{-}N}$ =18.5 Hz, C2), 148.2 (d, $^2J_{C\text{-}N}$ = 12.8 Hz, C5), 110.5 (C6). 85.4 (t, J = 42.0 Hz, C4'), 85.0 (dd, J = 36.4 Hz, $^1J_{C\text{-}N}$ = 10.4 Hz C1'), 67.3 (t, J = 39.1 Hz, C3'), 59.3 (d, J = 37.0 Hz, C5'), 39.7 (t, J = 37.0 Hz, C2'), 17.5-16.7 (8C, $C_{b\text{-}i}$), 13.4 (C_a)**, 12.9 (C_d)**, 12.8 (C_j)**, 12.4 (C_g)**.

^{15}N-NMR (41 MHz, $CDCl_3$): δ (ppm) = -228.7, -228.7.

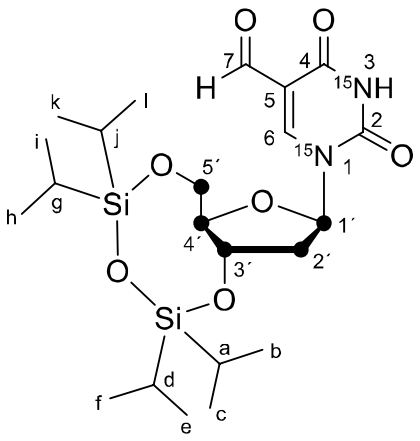
HRMS (ESI): berechnet für $C_{16}{}^{13}C_5H_{38}I^{15}N_2O_6Si_2^+$ [(M+H)$^+$]: 604.1416, gefunden 604.1417.

5.2.2.6 3',5'-(1,1,3,3,-Tetraisopropyldisiloxan-1,3-diyl)-5-formyl-2'-desoxy-(1',2',3',4',5'-$^{13}C_5$,*N*^{1},*N*^{3}-$^{15}N_2$)-uridin (10)



10

Die Reaktion wurde nach einer modifizierten Literaturvorschrift durchgeführt.[184] Analog zur ALLGEMEINEN REAKTIONSVORSCHRIFT II wurden in einem Druckreaktor Pd_2dba_3•$CHCl_3$ (77.0 mg, 74.6 µmol, 0.1 Äq.), PPh_3 (0.12 g, 0.45 mmol, 0.6 Äq.) und **9** (0.45 g, 0.75 mmol, 1.0 Äq.) vorgelegt und in Toluol (15 mL) gelöst. Das Reaktionsgefäß wurde unter Rühren mit CO (3.5 bar) gefüllt um die Versiegelung des Aufbaus zu prüfen (3.5 bar CO für 15 min). Der Vorgang wurde zur Sättigung der Reaktionslösung drei mal wiederholt. Die Reaktionsmischung wurde auf 60 C erwärmt und eine Lösung von $HSnBu_3$ (0.26 g, 0.24mL, 0.89 mmol, 1.2 Äq.) in Toluol (0.3 mL) mittels einer Spritzenpumpe über 15 h bei 65 °C und 3.5 bar CO zugetropft (40 µL/h). Anschließend wurde die Reaktionslösung über SiO_2 filtriert, mit EtOAc (100 mL) gespült und alle flüchtigen Bestandteile wurden unter vermindertem Druck entfernt. Säulenchromatographische Aufreinigung (SiO_2, 10 w% K_2CO_3, *i*Hexan/EtOAc = 3:1) ergab das formylierte Nukleosid **10** (260 mg, 0.51 mmol, 69%) in Form eines gelblichen Schaums.



R_f = 0.37 (SiO_2, *i*Hexan/EtOAc = 2:1).

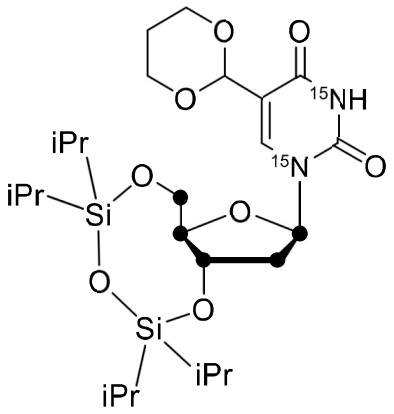
^{1}H NMR (600 MHz, $CDCl_3$): δ/ppm = 9.98 (s, 1H, C7-H), 8.48 (s, 1H, C6-H), 8.14 (d, $^1J_{N-H}$ = 91.2 Hz, 1H, N3-H), 6.01 (d, $^1J_{C-H}$ = 174.5 Hz, ^3J = 7.5 Hz 1H, C1'-H), 4.47 (dd, $^1J_{C-H}$ = 145.2 Hz, ^3J = 8.1 Hz, 1H, C3'-H), 4.17 (ddd, $^1J_{C-H}$ = 140.9 Hz, ^3J = 13.0 Hz, ^3J = 4.2 Hz 1H, C5'-H_a)*, 4.03 (dd, $^1J_{C-H}$ = 140.0 Hz, ^2J = 13.0 Hz 1H, C5'-H_b)*, 3.84 (d, $^1J_{C-H}$ = 146.5 Hz ,1H, C4')*, 2.53 (d, $^1J_{C-H}$ = 136.7 Hz, 1H, C2'-H_a)**, 2.34 (d, $^1J_{C-H}$ = 136.7 Hz, 1H, C2'-H_b)**, 1.10-1.01 (m, 28H, C_{a-l}-H).

^{13}C-NMR (126 MHz, $CDCl_3$): δ/ppm = 185.3 (C7), 161.0 (d, $^1J_{C-N}$ = 9.5 Hz, C4), 145.3 (C2), 145.2 (C6), 110.8 (C5), 86.0 (C4'), 85.5 (dd, J = 36.2 Hz, $^1J_{C-N}$ = 9.8 Hz C1'), 67.6 (t, J = 39.9 Hz, C3'), 60.3 (d, J = 43.9 Hz, C5'), 40.0 (t, J = 36.0 Hz, C2'), 17.4 (C_b)*, 17.3 (C_c)*, 17.2 (C_e)*, 17.2 (C_f)*, 17.1 (C_h)*, 17.0 (C_i)*, 17.0 (C_k)*, 16.8 (C_l)*, 13.5 (C_a)**, 13.0 (C_d)**, 12.7 (C_j)**, 12.5 (C_g)**.

^{15}N-NMR (41 MHz, $CDCl_3$): δ (ppm) = -224.8, -224.9.

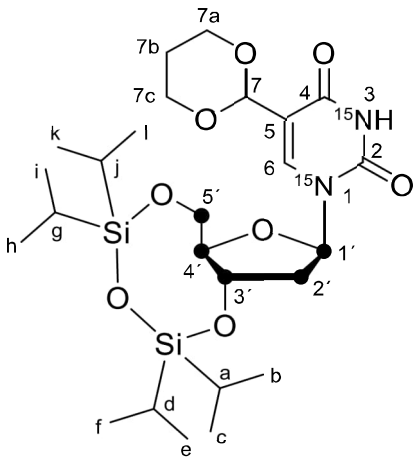
HRMS (ESI): berechnet für $C_{17}{}^{13}C_5H_{39}{}^{15}N_2O_7Si_2^+$ [$(M+H)^+$]: 506.2399, gefunden 506.2401.

5.2.2.7 3',5'-(1,1,3,3,-Tetraisopropyldisiloxan-1,3-diyl)-(1,3-dioxan-2-yl)-2'-desoxy-(1',2',3',4',5'-$^{13}C_5$,*N*^{1},*N*^{3}-$^{15}N_2$)-uridin (11)



11

Nukelosid **10** (0.26 g, 0.54 mmol, 1.0 Äq.) wurde in CH_2Cl_2 (10 mL) gelöst, auf 0 °C abgekühlt und mit 1,3-Propanediol 150 mg, 140 µL, 2.05 mmol, 4.0 Äq.), Triethylorthoformiat (150 mg, 170 µL, 1.03 mmol, 2.0 Äq.) and *p*-TSA (20.0 mg, 0.103 mmol, 0.2 Äq.) umgesetzt. Die Reaktionmsichung wurde 30 min bei 0 °C und 24 h bei Raumtemperatur gerührt. Die Reaktion wurde durch Zugabe von H_2O (20 mL) abgebrochen und mit CH_2Cl_2 (4 x 30 mL). extrahiert. Verunigrten organischen Phasen wurde mit einer gesättigten, wässrigen NaCl-Lösung (2 x 80 mL) gewaschen und über Na_2SO_4 getrocknet. Aufeinigung mittels Säulenchromatographie (SiO_2, *i*Hexan/EtOAc = 2:1) ergab die Verbindung **11** (140 mg, 0.25 mmol, 48%) in Form eines farblosen Schaums.



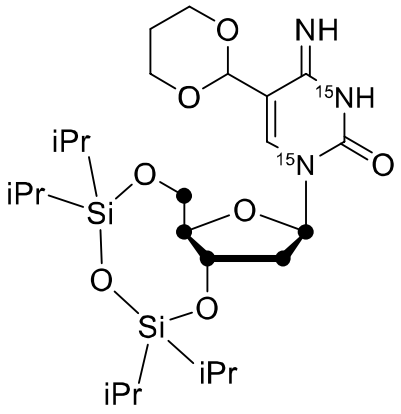
R_f = 0.29 (SiO_2, *i*Hexan/EtOAc = 1:1).

^{1}H NMR (600 MHz, $CDCl_3$): δ/ppm = 8.15 (d, $^{1}J_{N-H}$ = 91.8 Hz, 1H, N3-H), 7.71 (d, 1H, $^{1}J_{N-H}$ = 2.2 Hz, C6-H), 6.10 (dd, $^{1}J_{C-H}$ = 171.3 Hz, ^{3}J = 6.5 Hz 1H, C1'-H), 5.52 (s, 1H, C7-H), 4.57 (dd, $^{1}J_{C-H}$ = 144.4 Hz, ^{3}J = 8.2 Hz, 1H, C3'-H), 4.24-4.18 (m 2H, $C7_c$-H_a, C5'-H_a)*, 4.01-3.93 (m, 3H, $C7_c$-H_b, $C7_a$-H_a, C5'-H_b)*, 3.81 (d, $^{1}J_{C-H}$ = 140.05 Hz ,1H, C4'), 3.76 (m, 1H, $C7_a$-H_b)*, 2.51 (d, $^{1}J_{C-H}$ = 135.5 Hz, 1H, C2'-H_a)**, 2.33 (d, $^{1}J_{C-H}$ = 135.5 Hz, 1H, C2'-H_b)**, 2.19-2.12 (m, 1H, $C7_b$-H_a)***, 1.96-1.95 (m, 1H, $C7_b$-H_b)***, 1.15-1.04 (m, 28H, C_{a-l}-H).

^{13}C-NMR (126 MHz, $CDCl_3$): δ/ppm = 161.1(d, $^{1}J_{C-N}$ = 9.7 Hz, C4), 149.6 (t, $^{1}J_{C-N}$ = 9.7 Hz, C2), 138.9 (d, $^{1}J_{C-N}$ = 12.9 Hz, C6), 112.7 (d, $^{1}J_{C-N}$ = 6.69 Hz, C5), 95.4 (C7), 85.2 (t, J = 40.5 Hz, C4'), 85.5 (dd, J = 36.2 Hz, $^{1}J_{C-N}$ = 11.4 Hz C1'), 70.2 (t, J = 39.1 Hz, C3'), 61.9 (d, J = 44.5 Hz, C5'), 40.0 (t, J = 38.3 Hz, C2'), 31.6 ($C7_a$)*, 25.7 ($C7_c$)*, 17.6 (C_b)/*, 17.6 (2C, $C_{c,e}$)**, 17.5 (C_f)**, 17.4 (C_h)**, 17.2 (C_i)**, 17.1 (C_k)**, 17.0 (C_l)**, 14.4 ($C7_b$), 13.5 (C_a)***, 13.3 (C_d)***, 12.9 (C_j)***, 12.7 (C_g)***.

^{15}N-NMR (41 MHz, $CDCl_3$): δ (ppm) = -225.5, -225.6.

5.2.2.8 3',5'-(1,1,3,3,-Tetraisopropyldisiloxan-1,3-diyl)-(1,3-dioxan-2-yl)-2'-desoxy-(1',2',3',4',5'-$^{13}C_5$,N^1,N^3-$^{15}N_2$)-cytidin (12)

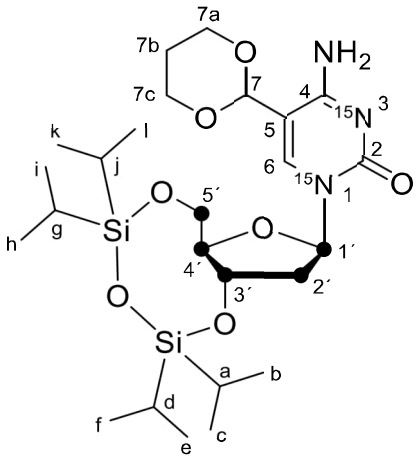


12

Verbidnung **11** (140 mg, 0.25 mmol, 1.0 Äq.) wurde in CH_2Cl_2 (2.5 mL) gelöst, auf 0 °C gekühlt und und mit DMAP (6.0 mg, 49.7 µmol, 0.2 Äq.) und NEt_3 (0.34 mL, 0.25 g, 2.48 mmol, 10.0 Äq.) versetzt. Bei 0 °C wurde TPS-Cl (110 mg, 0.37 mmol, 1.5 Äq.) portionsweise über einen Zeitraum von 30 min zugegeben. Die Reaktionsmischung wurde auf Raumtemperatur erwärmen lassen und für 16 h bei Raumtemperatur gerührt. Die Reaktion wurde durch Zugabe einer gesättigten wässrigen NH_4Cl-Lösung (10 mL) neutralisiert, und mit CH_2Cl_2 (3 x 30 mL)

extrahiert. Die vereinigten organischen Phasen wurden über Na_2SO_4 getrocknet und die Lösungsmittel wurden *in vacuo* entfernt. Das erhaltene Rohprodukt wurde ohne weitere Aufreinigungsschritte weiter umgesetzt.

Das erhaltene Rohprodukt (0.25 mmol, 1.0 Äq.) wurde in 1,4-Dioxan (5 mL) gelöst und tropfenweise mit einer wässrigen Lösung von NH_4OH (25% aq., 1.5 mL) versetzt und für 20 h bei Raumtemperatur gerührt. Die Reaktion wurde durch Zugabe einer gesättigten wässrigen NH_4Cl-Lösung (20 mL) beendet und mit EtOAc (3 x 30 mL) extrahiert. Die vereinigten organischen Phasen wurden mit einer gesättigten wässrigen NaCl-Lösung (100 mL) gewaschen und anschließend über Na_2SO_4 getrocknet. Lösungsmittel wurden *in vacuo* entfernt und das Rohprudukt säulenchromatographisch aufgereinigt (SiO_2, CH_2Cl_2/MeOH = 50:1) um die Verbindung **12** in Form eines gelblichen Schaums (92 mg, 0.16 mmol, 66%) zu erhalten.



R_f = 0.44 (SiO_2, CH_2Cl_2/MeOH = 10:1).

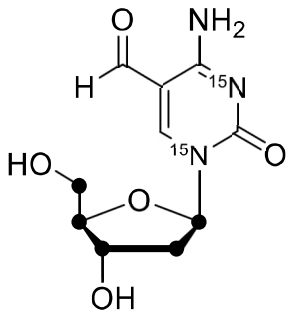
^{1}H NMR (600 MHz, $CDCl_3$): δ/ppm = 7.81 (d, 1H, $^1J_{N\text{-}H}$ = 2.5 Hz, C6-H), 6.03 (dd, $^1J_{C\text{-}H}$ = 171.1 Hz, ^3J = 6.4 Hz 1H, C1'-H), 5.26 (s, 1H, C7-H), 4.52-3.60 (m, 8H, C3'-H, C4'-H, C5'-H, $C7_{a\text{-}c}$), 2.53 (d, $^1J_{C\text{-}H}$ = 122.5 Hz, 1H, C2'-H_a)*, 2.29 (d, $^1J_{C\text{-}H}$ = 122.5 Hz, 1H, C2'-H_b)*, 2.24-2.13 (m, 1H, $C7_b$-H_a)**, 1.49-1.45 (m, 1H, $C7_b$-H_b)**, 1.12-0.90 (m, 28H, $C_{a\text{-}l}$-H).

^{13}C-NMR (126 MHz, $CDCl_3$): δ/ppm = 163.3 (d, $^1J_{C\text{-}N}$ = 5.9 Hz, C2)*, 163.3 (d, $^1J_{C\text{-}N}$ = 6.0 Hz, C4)*, 140.2 (t, $^1J_{C\text{-}N}$ = 13.9 Hz, C6), 103.4 (C5), 100.1 (C7), 85.2-84.6 (m, 2C, C1', C4'), 67.3 (t, J = 38.3 Hz, C3'), 60.2 (d, J = 43.5 Hz, C5'), 40.1 (t, J = 37.3 Hz, C2'), 25.6 ($C7_{a,c}$) 17.5 (C_b)*, 17.5 (2C, C_c)**, 17.5 (C_e)**, 17.3 (C_f)**, 17.1 (C_h)**, 17.0 ($C_{i,j}$)**, 16.9 (C_k)**, 13.4 (C_a)***, 13.0 (C_d)***, 12.7 (C_j)***, 12.4 (C_g)***.

^{15}N-NMR (41 MHz, $CDCl_3$): δ (ppm) = --206.7, -206.5.

HRMS (ESI): berechnet für $C_{21}{}^{13}C_5H_{46}N^{15}N_2O_7Si_2^+$ [(M+H)$^+$]: 563.2997, gefunden 563.2983.

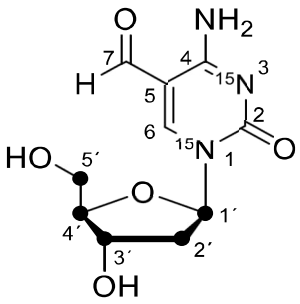
5.2.2.9 5-Formyl-2'-desoxy-(1',2',3',4',5'-$^{13}C_5$,N^1,N^3-$^{15}N_2$)-cytidin (1)



1

Verbindung **12** (92.0 mg, 0.16 mmol, 1.0 Äq.) wurde in THF (6 mL) gelöst und mit einer wässrigen HCl-Lösung (1 M, 0.65 mL, 4.0 Äq.) verseztz. Die Reaktionslösung wurde für 18 h bei Raumtemepratur gerührt, mit H_2O (50 mL) versetzt und mit EtOAc (3 x 20 mL) extrahiert. Die vereinigten organischen Phasen wurden mit einer gesättigten, wässrigen NaCl-Lösung (20 mL) gewaschen und über Na_2SO_4 getrocknet. Alle flüchtigen Bestandteile wurden *in vacuo* entfernt und das isolierte Nukleosid **12a** wurde ohne weitere Aufreinigung für die finale Entschützung verwendet.

Nach der modifizierten *ALLGEMEINEN REAKTIONSVORSCHRIFT I* wurde **12a** (74.0 mg, 0.15 mmol, 1.0 Äq.) in EtOAc (2.5 mL) mit HF • Pyridin (70%, 37.0 µL, 1.55 mmol, 10.0 Äq.) und Pyridin (61 µL) versetzt. Die Reaktionsmischung wurde 20 h bei Raumtemperatur gerührt, anschließend auf 0 °C gekühlt, mit TMS • OMe (0.7 mL, 4.93 mmol, 32.0 Äq.) versetzt und für 1 h bei 0 °C gerührt. Aufreinigung mittels semipräparativer HPL-Chromatographie (MeCN/H_2O, MeCN: 0 → 10%, 45 min, Retentionszeit: 25 min) ergab Verbindung **1** (13.0 mg, 49.2 µmol, 30%) in Form eines farblosen Pulvers.



Retentionszeit = 25 min (C_{18}-SiO_2, MeCN/H_2O, MeCN: 0 → 10%, 45 min).

^{1}H-NMR (400 MHz, D_2O): δ/ppm = 9.50 (s, 1H, C7-H), 8.84 (d, $^2J_{N-H}$ = 2.9 Hz, 1H, C6-H), 6.13 (dt, $^1J_{C-H}$ = 175.3 Hz, ^3J = 5.8 Hz 1H, C1'-H), 4.60 (d, $^1J_{C-H}$ = 151.9 Hbz, 1H,C3'-H), 4.30-3.93 (m, 2H, C4'-H + C5'-H), 3.74-3.52 (m, 1H, C5'-H), 2.76 (d, $^1J_{C-H}$ = 136.0 Hz, 1H, C2'-H), 2.51 (d, $^1J_{C-H}$ = 134.5 Hz, 1H, C2'-H).

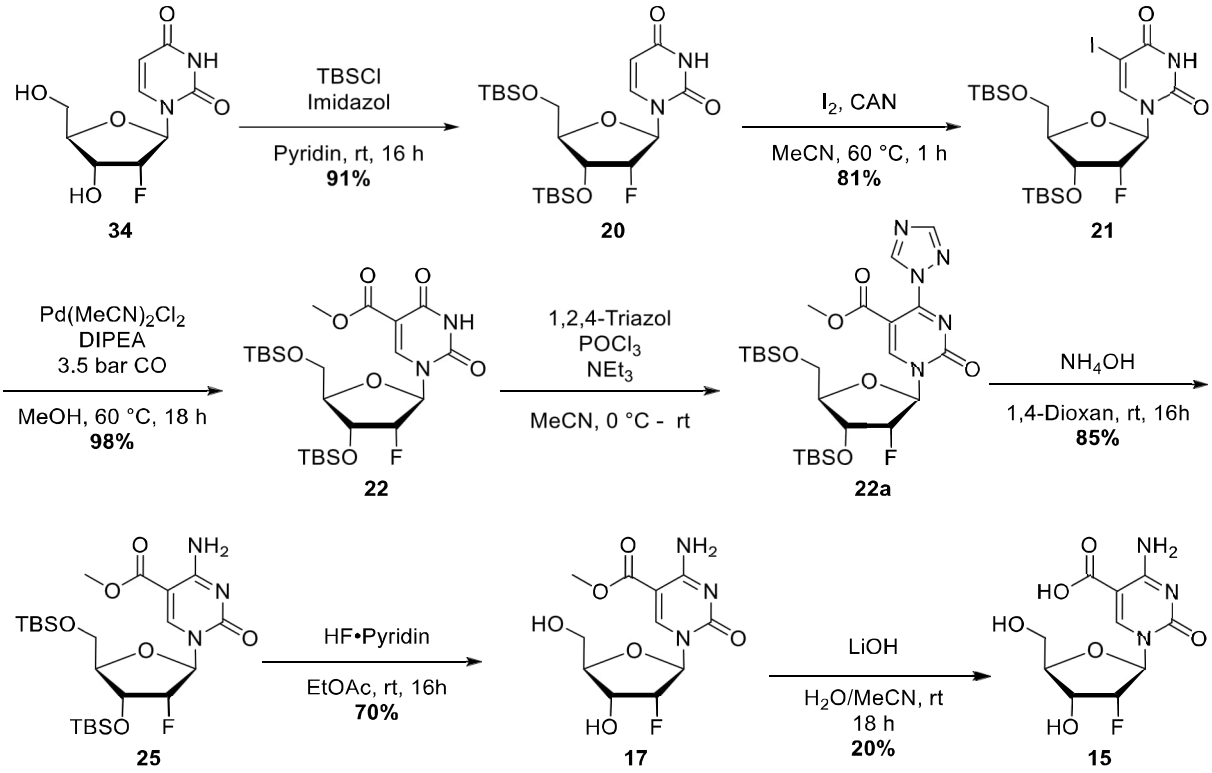
^{13}C-NMR (126 MHz, D_2O): δ/ppm = 190.4 (C-7), 161.9 (d, $^1J_{C-N}$ = 7.4 Hz, C-4), 154.8 (d, $^1J_{C-N}$ = 14.2 Hz, C-5), 154.2 (t, $^1J_{C-N}$ = 11.3 Hz, C-2), 105.3 (C-6), 87.8-86.7 (C-1' + C-4'), 69.8 (t, $^1J_{C-C}$ = 36.5, C-3'), 60.2 (d, J = 41.5 Hz, C-5'), 40.1 (t, $^1J_{C-C}$ = 36.0 Hz, C-2').

^{15}N-NMR (41 MHz, D_2O): δ/ppm = -209.9, -210.2.

HRMS (ESI): berechnet für $C_5{}^{13}C_5H_{14}N^{15}C_2O_5^+$ $[M+H]^+$ 263.1036, gemessen: 263.1040.

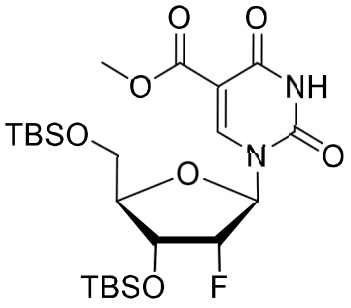
Die gemessenen Daten stimmen weitgehend mit der Literatur überein.[135]

5.2.3 Synthese von 5-Carboxy-2'-desoxy-2'-(*R*)-fluoro-cytidin (15)



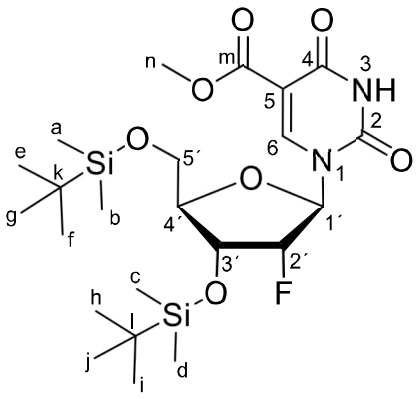
Schema 5.2: Syntheseschema von 2'-F-cadC (**15**) ausgehend von komerziell erwerblichem 2'-F-dU (**34**).

5.2.3.1 3',5'-Bis-*O*-(*tert*-butyldimethylsilyl)-5-(methoxycarbonyl)-2'-desoxy-2'-(*R*)-fluoro-uridin (22)



22

Das 5-iodierte Uridin-Derivat **21** wurde wie in der Literatur berichtet dargestellt.[177] Nach der *ALLGEMEINEN REAKTIONSVORSCHRIFT II* wurde das Nukelosid **21** (0.70 g, 1.17 mmol, 1.00 Äq.) mit $[Pd(MeCN)_2Cl_2]$ (15.1 mg, 58.0 µmol, 5.00 Mol%) und DIPEA (410 µL, 2.33 mmol, 2.00 Äq.) in Methanol (14 mL) umgesetzt und 18 h bei 65 °C und 3.5 bar CO gerührt. Aufeinigung mittel Säulenchromatographie (SiO_2, iHexan/EtOAc = 3:1) ergab die carboxylierte Verbindung **22** (0.61 mg, 1.14 mmol, 98%) in Form eines farblosen Feststoffes.



R_f = 0.37 (SiO_2, *i*Hexan/EtOAc = 2:1).

^{1}H-NMR (599 MHz, $CDCl_3$): δ/ppm = 8.42 (s, 1H, C6-H), 8.16 (s, 1H, N3-H), 6.03 (dd, $^3J_{H-F}$ = 14.8 Hz, ^3J = 4.0 Hz, 1H, C1'-H), 4.93 (dt, $^2J_{H-F}$ = 52.8 Hz, ^3J = 4.2 Hz, 1H, C2'-H), 4.34 (dt, $^3J_{H-F}$ = 11.5 Hz, ^3J = 5.0 Hz, 1H, C3'-H), 4.11-4.08 (m, 1H, C4'-H), 3.95 (dd, ^2J = 11.8 Hz, ^3J = 2.2 Hz, 1H, C5'-H_a)*, 3.86 (s, 1H, C_n-H_3), 3.77 (dd, ^2J = 11.8 Hz, ^3J = 3.0 Hz, 1H, C5'-H_b)*, 0.92 (s, 9H, C_{e-g}-H_3)**, 0.91 (s, 9H, C_{h-j}-H_3)**, 0.14-0.10 (m, 12H, C_{a-d}).

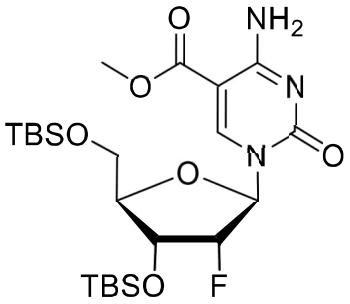
^{13}C-NMR (101 MHz, $CDCl_3$): δ/ppm = 163.0 (C_m), 158.4 (C4), 149.0 (C2), 147.7 (C6), 106.3 (C5), 92.3 (d, $^1J_{C–F}$ = 195.7 Hz, C2'), 88.9 (d, $^2J_{C–F}$ = 33.5 Hz, C1'), 85.7 (C4'), 70.0 (d, $^2J_{C–F}$ = 15.4 Hz, C3'), 62.0 (C5'), 52.6 (C_n), 26.1 (3C, C_{e-g})*, 25.8 (3C, C_{h-j})*, 18.6 (C_k)**, 18.3 (C_l)**, −4.6 (C_a)***, −5.0 (C_b)***, −5.3 (C_c)***, −5.4 (C_d)***.

^{19}F-NMR (377 MHz, $CDCl_3$): δ/ppm = −204.6 (dt, $^2J_{F–H}$ = 52.7 Hz, $^3J_{F–H}$ = 12.8 Hz, C2'-*F*).

HRMS (ESI): berechnet für $C_{23}H_{42}FN_2O_7Si_2$ $[M+H]^+$ 533.2509, gefunden: 533.2502.

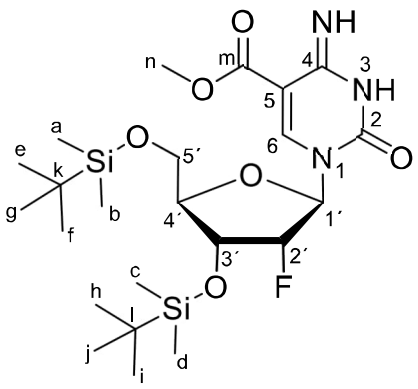
IR (ATR): ν/cm^{-1} = 3359 (w), 3253 (w), 2953 (w), 2929 (w), 2857 (w), 2892 (w), 1760 (m), 1713 (s), 1693 (s), 1614 (w), 1453 (m), 1390 (w), 1298 (w), 1254 (s), 1162 (m), 1122 (m), 1084 (s), 1069 (s), 1005 (w), 1004 (w), 984 (w), 938 (w), 910 (w), 869 (m), 830 (vs), 779 (vs), 758 (s), 672 (m).

5.2.3.2 3',5'-Bis-*O*-(*tert*-butyldimethylsilyl)-5-(methoxycarbonyl)-2'-desoxy-2'-(*R*)-fluoro-cytidin (25)



25

Nach der *ALLGEMEINEN REAKTIONSVORSCHRIFT III* wurde Verbindung **22** (0.48 g, 0.90 mmol, 1.0 Äq.) in MeCN (23.0 mL) mit 1,2,4-Triazol (0.56 g, 8.11 mmol, 9.0 Äq), $POCl_3$ (170 µL, 1.80 mmol, 2.0 Äq) und NEt_3 (1.11 mL, 8.02mmol, 8.9 Äq.) zum Triazolderivat (**22a**, R_f = 0.80 (SiO_2, *i*Hexan/EtOAc = 1:1) umgesetzt. Das erhaltene Triazolderivat wurde in 1,4-Dioxan (18 mL) mit einer wässrigen Lösung von NH_3 (25%, 5.7 mL) zur Reaktion gebracht. Säulenchromatographische Aufreinigung (SiO_2, *i*Hexan/EtOAc = 3:1 → *i*Hexan/EtOAc = 1:1) ergab Verbindung **25** in Form eines farblosen Schaumes (0.41 g, 0.77 mmol, 85%).



R_f = 0.11 (SiO_2, *i*Hexan/EtOAc = 2:1).

^{1}H-NMR (599 MHz, $CDCl_3$): δ/ppm = 8.51 (s, 1H, C6-H), 8.02 (d, ^3J = 4.8 Hz 1H, N3-H), 5.97 (d, ^3J = 4.8 Hz 1H, N4-H), 5.93 (dd, $^3J_{H–F}$ = 17.3 Hz, ^3J = 2.0 Hz, 1H, C1'-H), 4.98 (ddd, $^2J_{H–F}$ = 52.4 Hz, ^3J = 4.6 Hz, ^3J = 2.0 Hz, 1H, C2'-H), 4.21 (ddd, $^3J_{H–F}$ = 18.0 Hz, ^3J = 7.4 Hz, ^3J = 4.6 Hz 1H, C3'-H), 4.13-4.10 (m, 1H, C4'-H), 4.02 (dd, ^2J = 11.8 Hz, ^3J = 1.9 Hz, 1H, C5'-H_a)*, 3.84 (s, 1H, C_n-H_3), 3.80 (dd, ^2J = 11.8 Hz, ^3J = 3.6 Hz, 1H, C5'-H_b)*, 0.91 (s, 9H, C_{e-g}-H_3)**, 0.90 (s, 9H, C_{h-j}-H_3)**, 0.12-0.09 (m, 12H, C_{a-d}).

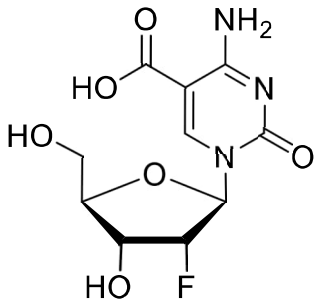
^{13}C-NMR (101 MHz, $CDCl_3$): δ/ppm = 165.3 (C_m), 163.8 (C4), 153.9 (C2), 147.5 (C6), 96.1 (C5), 92.7(d, $^1J_{C–F}$ = 192.8 Hz, C2'), 90.8 (d, $^2J_{C–F}$ = 34.6 Hz, C1'), 84.4 (C4'), 69.4 (d, $^2J_{C–F}$ = 16.4 Hz, C3'), 61.7 (C5'), 52.3 (C_n), 26.2 (3C, C_{e-g})*, 25.8 (3C, C_{h-j})*, 18.6 (C_k)**, 18.2 (C_l)**, −4.5 (C_a)***, −4.9 (C_b)***, −5.2 (C_c)***, −5.4 (C_d)***.

^{19}F-NMR (377 MHz, $CDCl_3$): δ/ppm = −200.8 (dt, $^2J_{F–H}$ = 52.5 Hz, $^3J_{F–H}$ = 17.6 Hz, C2'-*F*).

HRMS (ESI): berechnet für $C_{23}H_{43}FN_3O_6Si_2$ $[M+H]^+$ 533.2669, gefunden: 533.2659.

IR (ATR): v/cm^{-1} = 3255 (bw), 3104 (w), 2953 (w), 2930 (w), 2892 (w), 2858 (w), 1760 (m), 1713 (s), 1694 (s), 1642 (w), 1614 (m), 1453 (s), 1406 (w), 1391 (w), 1328 (w), 1298 (w), 1254 (s), 1230 (w), 1193 (w), 1162 (w), 1122 (m), 1084 (s), 1069 (s), 1005 (w), 983 (w), 938 (w), 910 (w), 863 (s), 840 (vs), 815 (s), 780 (vs), 758 (s), 672 (m).

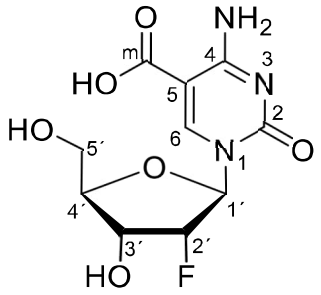
5.2.3.3 5-Carboxy-2'-desoxy-2'-(*R*)-fluoro-cytidin (15)



15

Nach der *ALLGEMEINEN REAKTIONSVORSCHRIFT I* wurde **25** (0.38 g, 0.72 mmol, 1.00 Äq.) in EtOAc (10.0 mL) mit HF • Pyridin (70%, 0.21 g, 0.27 mL, 10.7 mmol, 15.0 Äq.) versetzt. Die Reaktionsmischung wurde 20 h bei Raumtemperatur gerührt, anschließend auf 0 °C gekühlt, mit TMS • OMe (2.38 g, 3.20 mL, 22.8 mmol, 32.0 Äq.) versetzt und für 1 h bei 0 °C gerührt. Das erhaltene Nukleosid **17** wurde mittels semipräparativer HPL-Chromatographie aufgereinigt ($MeCN/H_2O$, MeCN: 0 → 80%, 45 min, Retentionszeit: 12 min) und ergab nach Lyophilisaton das Nukleosid **17** (43.5 mg, 0.14 mmol, 20%) in Form eines farblosen Pulvers.

Verbindung **17** (23.0 mg, 76.0 μmol, 1.00 Äq.) wurde in H_2O/MeCN = 1:1 (4 mL/4 mL) geslöst, mit LiOH (22.0 g, 0.91 mmol, 12.0 Äq.) versetzt und 24 h bei Raumtemperatur gerührt. Anschließend wurde die Reaktionslösung mit wässriger HCl (2 M, 10 mL) versetzt, auf pH = 3 angesäuert und alle flüchtigen Bestandteile wurdde *in vacuo* entfernt. Aufreinigung mittels semipräparativer HPL-Chromatographie ($MeCN/H_2O$, MeCN: 0 → 20%, 45 min, Retentionszeit: 10 min) ergab Verbindung **15** (15.4 mg, 53.0 μmol, 70%) in Form eines farblosen Pulvers.



Retentionszeit = 10 min (C_{18}-SiO_2, $MeCN/H_2O$, MeCN: 0 → 20%, 45 min).

^{1}H-NMR (400 MHz, D_2O): δ/ppm = 8.51 (s, 1H, C6-H), 6.03 (d, $^{3}J_{H–F}$ = 19.2 Hz, 1H, C1'-H), 5.14 (dd, $^{2}J_{H–F}$ = 52.8 Hz, ^{3}J = 4.6 Hz, 1H, C2'-H), 4.35 (ddd, $^{3}J_{H–F}$ = 22.9 Hz, ^{3}J = 9.0 Hz, ^{3}J = 4.3 Hz, 1H, C3'-H), 4.16 (ddd, $^{3}J_{H–F}$ = 8.9 Hz, ^{3}J = 4.5 Hz, ^{3}J = 2.4 Hz 1H, C4'-H), 4.04 (dd, ^{2}J = 13.0 Hz, ^{3}J = 2.4 Hz, 1H, C5'-H_a)*, 3.85 (dd, ^{2}J = 13.0 Hz, ^{3}J = 4.5 Hz, 1H, C5'-H_b)*.

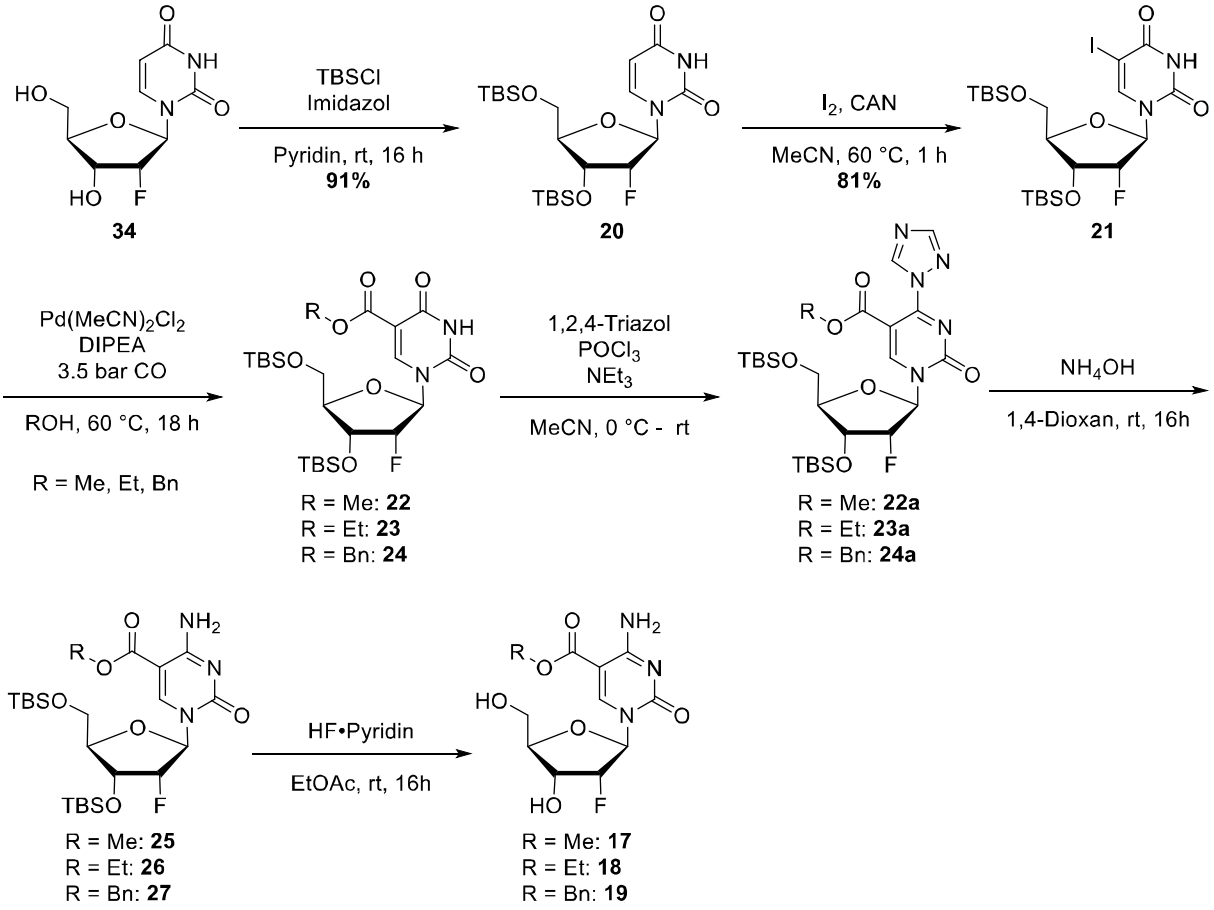
^{13}C-NMR (101 MHz, D_2O): δ/ppm = 170.3 (C_m), 164.6 (C4), 155.6 (C2), 146.9 (C6), 102.9 (C5), 91.6 (d, $^{1}J_{C-F}$ = 188.4 Hz, C2'), 89.9 (d, $^{2}J_{C-F}$ = 35.3 Hz, C1'), 82.3 (C4'), 67.8 (d, $^{2}J_{C-F}$ = 16.8 Hz, C3'), 59.7 (C5').

^{19}F-NMR (377 MHz, D_2O): δ/ppm = −200.7 (ddd, $^{2}J_{F–H}$ = 52.8 Hz, $^{3}J_{F–H}$ = 22.8 Hz, $^{3}J_{F–H}$ = 19.2 Hz, C2'-*F*).

HRMS (ESI): berechnet für $C_{10}H_{13}FN_3O_6$ $[M+H]^+$ 290.0783, gefunden: 290.0786.

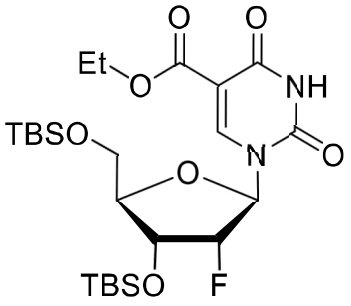
IR (ATR): v/cm^{-1} = 3334 (m), 3206 (m), 2928 (w), 2892 (w), 2857 (w), 2227 (w), 1647 (vs), 1493 (s), 1408 (m), 1361 (m), 1295 (m), 1252 (m), 1097 (s), 1060 (s), 986 (w), 950 (w), 917 (w), 833 (m), 813 (w), 781 (m), 703 (w), 672 (w).

5.2.4 Synthese von 5-Meth-/Eth-/Benzyl-oxy-2'-desoxy-2'-(*R*)-fluoro-cytidin (17, 18, 19)



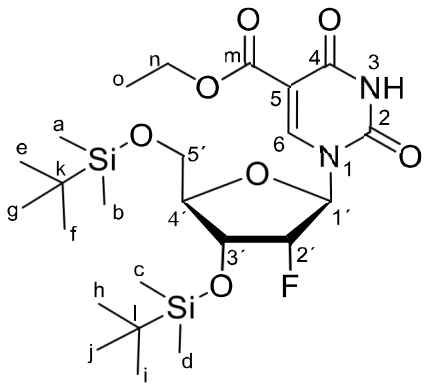
Schema 5.3: Syntheseschema von 2'-F-cadC *prodrugs*, 5-CO_2Me-2'-F-dC (**17**), 5-CO_2Et-2'-F-dC (**18**) und 5-CO_2Bn-2'-F-dC (**19**) ausgehend von komerziell erwerblichem 2'-F-dU (**34**).

5.2.4.1 3',5'-Bis-*O*-(*tert*-butyldimethylsilyl)-5-(ethoxycarbonyl)-2'-desoxy-2'-(*R*)-fluoro-uridin (23)



23

Nach der *ALLGEMEINEN REAKTIONSVORSCHRIFT II* wurde das Nukelosid **21** (0.50 g, 0.83 mmol, 1.00 Äq.) mit $[Pd(MeCN)_2Cl_2]$ (11.0 mg, 42.0 µmol, 5.00 Mol%) und DIPEA (0.22 g, 0.29 mL, 1.66 mmol, 2.00 Äq.) in EtOH (10 mL) umgesetzt und 18 h bei 65 °C und 3.5 bar CO gerührt. Aufeinigung mittel Säulenchromatographie (SiO_2, *i*Hexan/EtOAc = 3:1) ergab die carboxylierte Verbindung **23**(0.44 g, 0.80 mmol, 96 %) in Form eines farblosen Feststoffes.



R_f = 0.40 (SiO_2, *i*Hexan/EtOAc = 3:1).

^{1}H-NMR (599 MHz, $CDCl_3$): δ/ppm = 8.37 (s, 1H, C6-H), 8.25 (s, 1H, N3-H), 5.99c (dd, $^3J_{H-F}$ = 15.4 Hz, ^3J = 3.7 Hz, 1H, C1'-H), 4.96 (ddd, $^2J_{H-F}$ = 52.7 Hz, ^3J = 8.7 Hz, ^3J = 1.3 Hz 1H, C2'-H), 4.38-4.31 (m, 3H, C3'-H, C_n-H_2), 4.10-4.06 (m, 1H, C4'-H), 3.95 (dd, ^2J = 11.8 Hz, ^3J = 2.3 Hz, 1H, C5'-H_a)*, 3.77 (dd, ^2J = 11.8 Hz, ^3J = 3.6 Hz, 1H, C5'-H_b)*, 1.35 (t, ^3J = 7.1 Hz, 1H, C_o-H_3), 0.91 (s, 9H, C_{e-g}-H_3)**, 0.90 (s, 9H, C_{h-j}-H_3)**, 0.13-0.10 (m, 12H, C_{a-d}).

^{13}C-NMR (101 MHz, $CDCl_3$): δ/ppm = 162.6 (C_m), 158.5 (C4), 149.1 (C2), 147.7 (C6), 106.7 (C5), 92.3 (d, $^1J_{C-F}$ = 194.9 Hz, C2'), 89.3 (d, $^2J_{C-F}$ = 33.8 Hz, C1'), 85.5 (C4'), 70.0z (d, $^2J_{C-}$

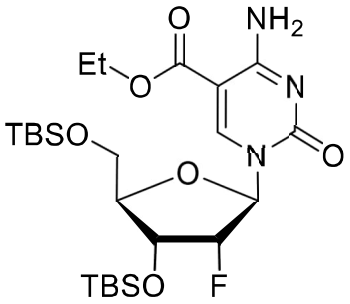
$_F$ = 15.4 Hz, C3'), 62.1 (C5'), 61.7 (C_n), 26.1 (3C, C_{e-g})*, 25.8 (3C, C_{h-j})*, 18.6 (C_k)**, 18.3 (C_l)**,14.5 (C_o), −4.6 (C_a)***, −-4.9, (C_b)***, −5.3 (C_c)***, −5.4 (C_d)***.

^{19}F-NMR (377 MHz, $CDCl_3$): δ/ppm = −206.6 (ddd, $^2J_{F-H}$ = 53.5 Hz, $^3J_{F-H}$ = 13.4 Hz, $^3J_{F-H}$ = 2.4 Hz, C2'-*F*).

HRMS (ESI): berechnet für $C_{24}H_{44}FN_2O_7Si_2$ $[M+H]^+$ 547.2666, gefunden: 547.2666.

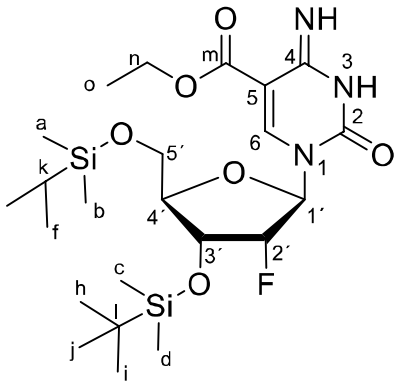
IR (ATR): v/cm^{-1} = 3206 (w), 3069 (w), 2954 (w), 2929 (w), 2887 (w), 2857 (w), 1753 (m), 1712 (s), 1616 (m), 1461 (m), 1438 (w), 1389 (w), 1360 (w), 1300 (w), 1257 (s), 1228 (w), 1187 (w), 1167 (w), 1131 (m), 1072 (vs), 1005 (w), 983 (w), 938 (w9; 905 (w), 860 (m), 834 (vs), 814 (m), 778 (vs), 704 (w9; 671 (m).

5.2.4.2 3',5'-Bis-*O*-(*tert*-butyldimethylsilyl)-5-(ethoxycarbonyl)-2'-desoxy-2'-(*R*)-fluoro-cytidin (26)



26

Nach der *ALLGEMEINEN REAKTIONSVORSCHRIFT III* wurde Verbindung **23** (0.10 g, 0.18 mmol, 1.0 Äq.) in MeCN (5.0 mL) mit 1,2,4-Triazol (0.11 g, 1.65 mmol, 9.0 Äq), $POCl_3$ (0.06 g, 0.34 µL, 0.37 mmol, 2.0 Äq) und NEt_3 (0.16 g, 0.22 mL, 1.57 mmol, 8.6 Äq.) zum Triazolderivat (**23a,** R_f = 0.70 (SiO_2, *i*Hexan/EtOAc = 1:1) umgesetzt. Das erhaltene Triazolderivat **23a** wurde in 1,4-Dioxan (5.0 mL) mit einer wässrigen Lösung von NH_3 (25%, 1.2 mL) zur Reaktion gebracht. Säulenchromatographische Aufreinigung (SiO_2, *i*Hexan/EtOAc = 1:1) ergab Verbindung **26** in Form eines farblosen Schaumes (62.0 mg, 0.11 mmol, 62%).



R_f = 0.50 (SiO_2, *i*Hexan/EtOAc = 1:1).

^{1}H-NMR (599 MHz, $CDCl_3$): δ/ppm = 8.43 (s, 1H, C6-H), 8.00 (d, ^4J = 4.5 Hz 1H, N3-H), 6.07 (d, ^4J = 3.7 Hz 1H, C4-NH)*, 5.86 (dd, $^3J_{H–F}$ = 18.0 Hz, ^3J = 2.0 Hz, 1H, C1'-H), 5.05 (ddd, $^2J_{H–F}$ = 52.9 Hz, ^3J = 4.7 Hz, ^3J = 2.1 Hz, 1H, C2'-H), 4.42-4.21 (m, 3H, C3'-H, C_n-H_3), 4.13-4.08 (m, 1H, C4'-H), 3.99 (dd, ^2J = 11.8 Hz, ^3J = 2.1 Hz, 1H, C5'-H_a)**, 3.80 (dd, ^2J = 11.8 Hz, ^3J = 4.6 Hz, 1H, C5'-H_b)**, 1.35 (t, ^3J = 7.2 Hz, 1H, C_o-H_3), 0.91 (s, 18H, C_{e-g}-H_3, C_{h-j}-H_3), 0.12-0.09 (m, 12H, C_{a-d}).

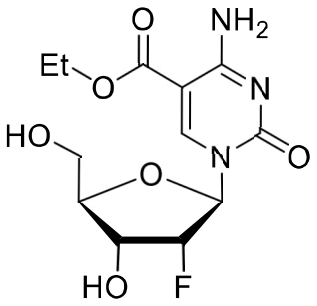
^{13}C-NMR (101 MHz, $CDCl_3$): δ/ppm = 164.7 (C_m), 163.8 (C4), 153.8 (C2), 148.9 (C6), 96.4 (C5), 92.6 (d, $^1J_{C–F}$ = 192.2 Hz, C2'), 90.8 (d, $^2J_{C–F}$ = 35.2 Hz, C1'), 84.5 (C4'), 69.8 (d, $^2J_{C–F}$ = 18.2 Hz, C3'), 62.4 (C5'), 61.5 (C_n), 26.1 (3C, C_{e-g})*, 25.8 (3C, C_{h-j})*, 18.6 (C_k)**, 18.2 (C_l)**, 14.5 (C_o), −4.5 (C_a)***, −4.9 (C_b)***, −5.1 (C_c)***, −5.3 (C_d)***.

^{19}F-NMR (377 MHz, $CDCl_3$): δ/ppm = −200.1 (dt, $^2J_{F–H}$ = 52.7 Hz, $^3J_{F–H}$ = 18.1 Hz, C2'-*F*).

HRMS (ESI): berechnet $C_{24}H_{45}FN_3O_6Si_2$ $[M+H]^+$ 546.2825, gefunden: 546.2829.

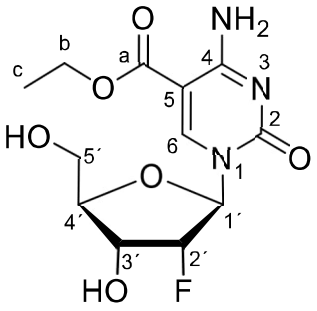
IR (ATR): v/cm^{-1} = 3206 (w), 3069 (w), 2954 (w), 2929 (w), 2888 (w), 2857 (w), 1754 (s), 1712 (s), 1617 (m), 1461 (s), 1405 (w), 1361 (w), 13007 (m), 1257 (s), 1167 (m), 1131 (s), 1072 (vs), 1005 (w), 860 (m), 840 (vs), 778 (vs), 672 (m).

5.2.4.3 5-Ethoxy-2'-desoxy-2'-(*R*)-fluoro-cytidin (18)



18

Nach der *ALLGEMEINEN REAKTIONSVORSCHRIFT I* wurde **26** (62.0 mg, 0.11 mmol, 1.0 Äq.) in EtOAc (1.7 mL) mit HF • Pyridin (70%, 44.0 µL, 1.70 mmol, 15.0 Äq.) versetzt. Die Reaktionsmischung wurde 20 h bei Raumtemperatur gerührt, anschließend auf 0 °C gekühlt, mit TMS • OMe (0.38 g, 0.50 mL, 3.64 mmol, 32.0 Äq.) versetzt und für 1 h bei 0 °C gerührt. Aufreinigung mittels semipräparativer HPL-Chromatographie (MeCN/H_2O, MeCN: 0 → 40%, 45 min, Retentionszeit: 26 min) ergab Verbindung **18** (20.0 mg, 63.0 µmol, 57%) in Form eines farblosen Pulvers.



Retentionszeit = 26 min (C_{18}-SiO_2, MeCN/H_2O, MeCN: 0 → 40%, 45 min).

^{1}H-NMR (400 MHz, D_2O): δ/ppm = = 9.21 (s, 1H, C6-H), 6.19 (dd, $^3J_{H–F}$ = 17.2 Hz, 1H, C1'-H), 5.21 (dd, $^2J_{H–F}$ = 52.4 Hz, ^3J = 4.1 Hz, 1H, C2'-H), 4.50-4.42 (m, 3H, C3'-H, C_b-H_2), 4.31 (dt, ^3J = 9.5 Hz, ^3J = 2.6 Hz 1H, C4'-H), 4.22 (dd, ^2J = 13.1 Hz, ^3J = 2.5 Hz, 1H, C5'-H_a)*, 3.98 (dd, ^2J = 13.1 Hz, ^3J = 2.6 Hz, 1H, C5'-H_b)*1.47 (t, ^3J = 7.1 Hz, 3H, C_a-H_3).

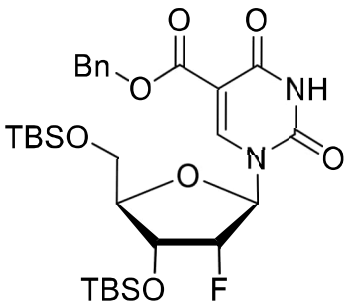
^{13}C-NMR (101 MHz, D_2O): δ/ppm = 167.1 (C_a), 166.3 (C4), 156.3 (C2), 150.2 (C6), 97.8 (C5), 94.7 (d, $^1J_{C–F}$ = 185.3 Hz, C2'), 90.3 (d, $^2J_{C–F}$ = 34.1 Hz, C1'), 83.0 (C4'), 67.2 (d, $^2J_{C–F}$ = 16.5 Hz, C3'), 61.4 (C_b), 59.2 (C5'), 14.1 (C_c).

^{19}F-NMR (377 MHz, D_2O): δ/ppm = −201.7 (ddd, $^2J_{F-H}$ = 53.2 Hz, $^3J_{F-H}$ = 17.1 Hz, $^3J_{F-H}$ = 7.9 Hz, C2'-*F*).

HRMS (ESI): berechnet für $C_{12}H_{17}FN_3O_6$ $[M+H]^+$ 318.1096, gefunden: 318.1099.

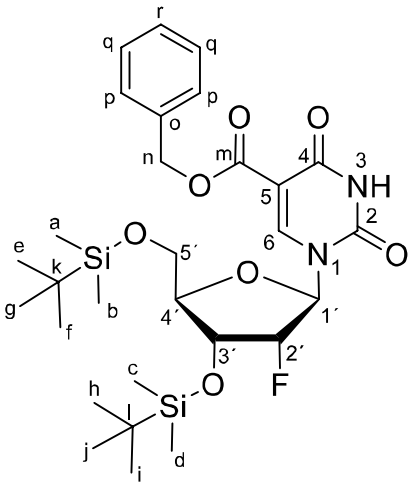
IR (ATR): v/cm^{-1} = 3420 (w), 3310 (w), 3069 (w), 2880 (w), 2863 (w), 1720 (s), 1605 (s), 1580 (m), 1405 (w), 1390 (m), 1330 (m), 1080 (s), 796 (m), 682 (m).

5.2.4.4 3',5'-Bis-*O*-(*tert*-butyldimethylsilyl)-5-(benzyloxycarbonyl)-2'-desoxy-2'-(*R*)-fluoro-uridin (24)



24

Nach der *ALLGEMEINEN REAKTIONSVORSCHRIFT II* wurde das Nukelosid **21** (0.40 g, 0.66 mmol, 1.0 Äq.) mit $[Pd(MeCN)_2Cl_2]$ (8.60 mg, 33.0 μmol, 5.0 Mol%), DIPEA (0.17 g, 0.23 mL, 1.33 mmol, 2.0 Äq.) und BnOH (0.14 g, 0.14 mL, 1.33 mmol, 2.0 Äq) in Toluol (10 mL) umgesetzt und 18 h bei 65 °C und 3.5 bar CO gerührt. Aufeinigung mittels Säulenchromatographie (SiO_2, *i*Hexan/EtOAc = 4:1 → *i*Hexan/EtOAc = 3:1) ergab die carboxylierte Verbindung **24** (0.15 g, 0.25 mmol, 37%) in Form eines farblosen Schaums.



R_f = 0.43 (SiO_2, *i*Hexan/EtOAc = 2:1).

^{1}H-NMR (599 MHz, $CDCl_3$): δ/ppm = 8.63 (s, 1H, N3-H), 8.41 (s, 1H, C6-H), 7.45-7.42 (m, 2H, C_p-H)*, 7.38-7.28 (m, 3H, C_q-H, C_r-H)*, 5.97 (dd, $^3J_{H–F}$ = 15.3 Hz, ^3J = 3.7 Hz, 1H, C1'-H), 5.33 (dd, ^2J = 14.2 Hz, ^3J = 3.1 Hz, 2H, C_n-H_2), 4.96 (ddd, $^2J_{H–F}$ = 52.9 Hz, ^3J = 4.0 Hz, ^3J = 1.1 Hz, 1H, C2'-H), 4.34 (dt, $^3J_{H–F}$ = 12.4 Hz, ^3J = 5.2 Hz 1H, C3'-H), 4.09-4.05 (m, 1H, C4'-H), 3.93 (dd, ^2J = 11.9 Hz, ^3J = 2.2 Hz, 1H, C5'-H_a)**, 3.75 (dd, ^2J = 11.9 Hz, ^3J = 3.3 Hz, 1H, C5'-H_b)**, 0.91 (s, 9H, C_{e-g}-H_3)***, 0.88 (s, 9H, C_{h-j}-H_3)***, 0.12 (s, 3H, C_a)****, 0.11 (s, 3H, C_b)****, 0.08 (s, 3H, C_c)****, 0.07 (s, 3H, C_d)****.

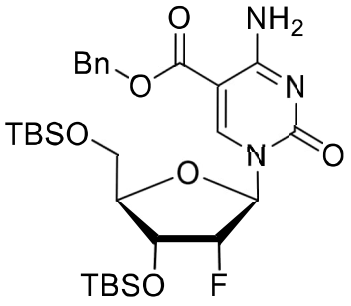
^{13}C-NMR (101 MHz, $CDCl_3$): δ/ppm = 162.5 (C_m), 158.6 (C4), 149.1 (C2), 148.1 (C6), 135.8 (C_o), 128.7 (C_p)*, 128.3 (C_q)*, 128.2 (C_r)*, 96.4 (C5), 92.3 (d, $^1J_{C–F}$ = 196.4 Hz, C2'), 89.5 (d, $^2J_{C–F}$ = 34.2 Hz, C1'), 85.5 (C4'), 70.0 (d, $^2J_{C–F}$ = 15.2 Hz, C3'), 67.1 (C_n), 62.1 (C5'), 26.1 (3C, C_{e-g})**, 25.8 (3C, C_{h-j})**, 18.6 (C_k)***, 18.3 (C_l)***, −4.6 (C_a)****, −5.0 (C_b)****, −5.3 (C_c)****, −5.4 (C_d)****.

^{19}F-NMR (377 MHz, $CDCl_3$): δ/ppm = −203.7 (dt, $^2J_{F–H}$ = 53.8 Hz, $^3J_{F–H}$ = 14.2 Hz, C2'-*F*).

HRMS (ESI): berechnet für $C_{29}H_{46}FN_2O7_6Si_2$ $[M+H]^+$ 609.2822, gefunden: 609.2822.

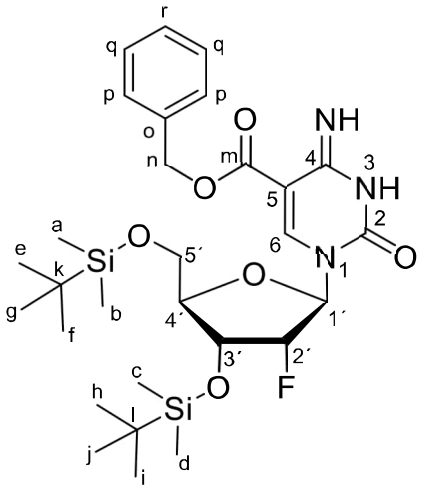
IR (ATR): ν/cm^{-1} = 3502 (w), 3067 (w), 2953 (w), 2928 (w), 2887 (w), 2856 (w), 1713 (s), 1692 (s), 1618 (m), 1498 (w), 1455 (m), 1390 (w), 1360 (w), 1296 (w), 1256 (s), 1226 (m), 1164 (m), 1127 (s), 1077 (s), 1004 (w), 981 (w), 938 (w), 905 (w), 859 (m), 830 (vs), 777 (vs), 734 (s), 695 (m), 671 (m).

5.2.4.5 3',5'-Bis-*O*-(*tert*-butyldimethylsilyl)-5-(benzyloxycarbonyl)-2'-desoxy-2'-(*R*)-fluoro-cytidin (27)



27

Nach der *ALLGEMEINEN REAKTIONSVORSCHRIFT III* wurde Verbindung **24** (0.20 g, 0.33 mmol, 1.0 Äq.) in MeCN (8.2 mL) mit 1,2,4-Triazol (0.20 g, 2.96 mmol, 9.0 Äq), $POCl_3$ (61.0 µL, 65.7 µmol, 2.0 Äq) und NEt_3 (0.39 mL, 2.82 mmol, 8.6 Äq.) zum Triazolderivat (**24a**, R_f = 0.80 (SiO_2, *i*Hexan/EtOAc = 1:1) umgesetzt. Das erhaltene Triazolderivat **24a** wurde in 1,4-Dioxan (6.5 mL) mit einer wässrigen Lösung von NH_3 (25%, 2.1 mL) zur Reaktion gebracht. Säulenchromatographische Aufreinigung (SiO_2, *i*Hexan/EtOAc = 3:1) ergab Verbindung **27** in Form eines farblosen Schaumes (38.0 mg, 62.5 µmol, 20%).



R_f = 0.40 (SiO_2, *i*Hexan/EtOAc = 1:1).

^{1}H-NMR (599 MHz, $CDCl_3$): δ/ppm = 8.45 (s, 1H, C6-H), 7.97 (s, 1H, N3-H)*, 7.41-7.32 (m, 5H, $C_{p\text{-}r}$-H), 6.95 (s, 1H, C4-NH)*, 5.82 (dd, $^3J_{H-F}$ = 18.7 Hz, ^3J = 2.0 Hz, 1H, C1'-H), 5.35 (d, ^2J = 12.4 Hz, 2H, C_n-H_a)**, 5.23 (d, ^2J = 12.4 Hz, 2H, C_n-H_b)**, 5.05 (ddd, $^2J_{H-F}$ = 53.0 Hz, ^3J = 4.6 Hz, ^3J = 2.0 Hz, 1H, C2'-H), 4.24 (ddd, $^3J_{H-F}$ = 17.3 Hz, ^3J = 7.3 Hz, ^3J = 2.9 Hz, 1H,

C3'-H), 4.10-4.07 (m, 1H, C4'-H), 3.94 (dd, ^{2}J = 11.8 Hz, ^{3}J = 2.4 Hz, 1H, C5'-H_a)***, 3.75 (dd, ^{2}J = 11.8 Hz, ^{3}J = 4.7 Hz, 1H, C5'-H_b)***, 0.90 (s, 9H, $C_{e\text{-}g}$-H_3)****, 0.87 (s, 9H, $C_{h\text{-}j}$-H_3)****, 0.11 (s, 3H, C_a)*****, 0.09 (s, 3H, C_b)*****, 0.05 (s, 3H, C_c)*****, 0.03 (s, 3H, C_d)*****.

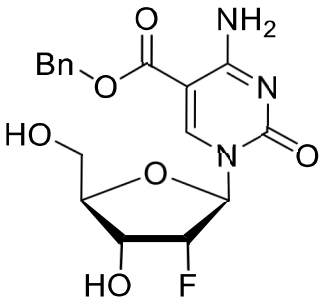
^{13}C-NMR (101 MHz, $CDCl_3$): δ/ppm = 164.6 (C_m), 163.8 (C4), 153.7 (C2), 149.3 (C6), 135.4 (C_o), 128.9 (C_p)*, 128.8 (C_q)*, 128.5 (C_r)*, 96.3 (C5), 92.6 (d, $^{1}J_{C–F}$ = 192.5 Hz, C2'), 92.0 (d, $^{2}J_{C–F}$ = 34.1 Hz, C1'), 84.6 (C4'), 69.9 (d, $^{2}J_{C–F}$ = 15.8 Hz, C3'), 67.0 (C_n), 62.5 (C5'), 26.1 (3C, $C_{e\text{-}g}$)**, 25.8 (3C, $C_{h\text{-}j}$)**, 18.6 (C_k)***, 18.2 (C_l)***, −4.5 (C_a)****, −4.9 (C_b)****, −5.2 (C_c)****, −5.3 (C_d)****.

^{19}F-NMR (377 MHz, $CDCl_3$): δ/ppm = −199.7 (dt, $^{2}J_{F–H}$ = 52.9 Hz, $^{3}J_{F–H}$ = 18.4 Hz, C2'-*F*).

HRMS (ESI): berechnet für $C_{29}H_{47}FN_3O_6Si_2$ $[M+H]^+$ 608.2982, gefunden: 608.2979.

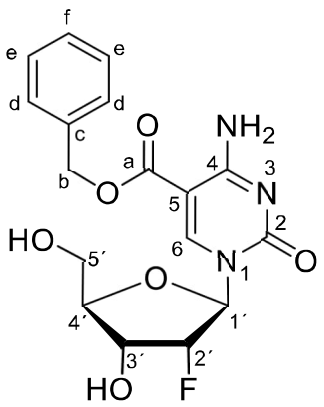
IR (ATR): ν/cm^{-1} = 3204 (w), 3067 (w), 2953 (w), 2929 (w), 2887 (w), 2856 (w), 1713 (s), 16912 (s), 1617 (m), 1498 (w), 1455 (m), 1391 (w), 1360 (w), 1296 (w), 1256 (s), 1226 (m), 1128 (s), 1077 (s), 1030 (w), 1004 (w), 982 (w), 938 (w), 905 (w), 859 (m), 837 (vs), 777 (vs), 734 (s), 695 (m), 671 (m).

5.2.4.6 5-Benzyloxy-2'-desoxy-2'-(*R*)-fluoro-cytidin (19)



19

Nach der *ALLGEMEINEN REAKTIONSVORSCHRIFT I* wurde **27** (38.0 mg, 62.5 µmol, 1.00 Äq.) in EtOAc (1.0 mL) mit HF • Pyridin (70%, 24.3 µL, 0.94 mmol, 15.0 Äq.) versetzt. Die Reaktionsmischung wurde 20 h bei Raumtemperatur gerührt, anschließend auf 0 °C gekühlt, mit TMS • OMe (0.21 g, 0.28 mL, 2.00 mmol, 32.0 Äq.) versetzt und für 1 h bei 0 °C gerührt. Aufreinigung mittels semipräparativer HPL-Chromatographie (MeCN/H_2O, MeCN: 0 → 80%, 45 min, Retentionszeit: 22 min) ergab Verbindung **19** (12.8 mg, 33.7 µmol, 54%) in Form eines farblosen Pulvers.



Retentionszeit = 22 min (C_{18}-SiO_2, MeCN/H_2O, MeCN: 0 → 80%, 45 min).

^{1}H-NMR (400 MHz, D_2O/CD_3CN 1:1): δ/ppm = 9.21 (s, 1H, C6-H), 7.67-7.61 (m, 5H. C_{d-f}-H) 6.18 (d, $^3J_{H–F}$ = 17.5 Hz, 1H, C1'-H), 5.5.56 (d, ^2J = 12.5 Hz, 2H, C_n-H_a)*, 5.51 (d, ^2J = 12.4 Hz, 2H, C_n-H_b)*, 5.25 (dd, $^2J_{H–F}$ = 52.3 Hz, ^3J = 3.7 Hz, 1H, C2'-H), 4.45 (ddd, 1H, $^3J_{H–F}$ = 24.2 Hz, ^3J = 9.2 Hz, ^3J = 4.3 Hz, C3'-H), 4.35-4.34 (m, 1H, C4'-H), 4.20 (dd, ^2J = 13.2 Hz, ^3J = 2.1 Hz, 1H, C5'-H_a)**, 3.97 (dd, ^2J = 13.2 Hz, ^3J = 2.9 Hz, 1H, C5'-H_b)**.

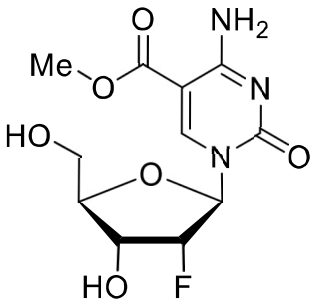
^{13}C-NMR (101 MHz, D_2O/CD_3CN 1:1): δ/ppm = 164.2 (C_a), 162.9 (C4), 154.3 (C2), 148.3 (C6), 134.9 (C_c), 128.1 (C_d)*, 127.9 (C_e)*, 127.4 (C_f)*, 96.0 (C5), 93.1 (d, $^1J_{C–F}$ = 183.2 Hz, C2'), 88.8 (d, $^2J_{C–F}$ = 38.2 Hz, C1'), 81.7 (C4'), 66.4 (d, $^2J_{C–F}$ = 17.5 Hz, C3'), 66.2 (C_b), 58.2 (C5').

^{19}F-NMR (377 MHz, D_2O/CD_3CN 1:1): δ/ppm = −200.1 (dt, $^2J_{F–H}$ = 53.2 Hz, $^3J_{F–H}$ = 18.2 Hz, C2'-*F*).

HRMS (ESI): berechnet für $C_{17}H_{19}FN_3O_6$ $[M+H]^+$ 380.1252, gefunden: 290.1251.

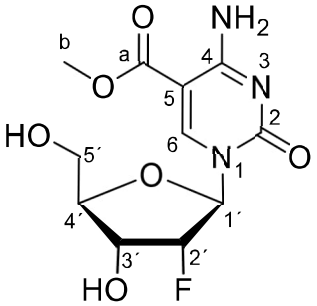
IR (ATR): v/cm^{-1} = 3312 (w), 3210 (w), 2839 (w), 2670 (w), 1620 (s), 1630 (s), 1456 (w), 1377 (w), 1322 (w), 1268 (w), 1187 (m), 1120 (s), 1030 (s), 1020 (m), 960 (w), 927 (w), 852 (w), 828 (vs), 791 (s), 723 (w), 662 (w).

5.2.4.7 5-Methoxy-2'-desoxy-2'-(*R*)-fluoro-cytidin (17)



17

Nach der *ALLGEMEINEN REAKTIONSVORSCHRIFT I* wurde **25** (0.61 g, 0.12 mmol, 1.00 Äq.) in EtOAc (2.0 mL) mit HF • Pyridin (70%, 41.0 µL, 1.73 mmol, 15.0 Äq.) versetzt. Die Reaktionsmischung wurde 17 h bei Raumtemperatur gerührt, anschließend auf 0 °C gekühlt, mit TMS • OMe (0.51 mL, 3.69 mmol, 32.0 Äq.) versetzt und für 1 h bei 0 °C gerührt. Aufreinigung mittels semipräparativer HPL-Chromatographie (MeCN/H_2O, MeCN: 0 → 80%, 45 min, Retentionszeit: 15 min) ergab Verbindung **17** (12.0 mg, 39.5 µmol, 34%) in Form eines farblosen Pulvers.



Retentionszeit = 15 min (C_{18}-SiO_2, MeCN/H_2O, MeCN: 0 → 80%, 45 min).

^{1}H-NMR (400 MHz, D_2O/CD_3CN 1:1): δ/ppm = 9.06 (s, 1H, C6-H), 5.90 (d, $^3J_{H–F}$ = 16.7 Hz, 1H, C1'-H), 4.93 (dd, $^2J_{H–F}$ = 52.5 Hz, ^3J = 4.0 Hz, 1H, C2'-H), 4.24-4.15 (m, 1H, C3'-H), 4.03 (d, ^3J = 9.3 Hz 1H, C4'-H), 3.96 (dd, ^2J = 13.0 Hz, ^3J = 2.2 Hz, 1H, C5'-H_a)*, 3.75-3.70 (m, 4H, C5'-H_b*, C_b-H_3).

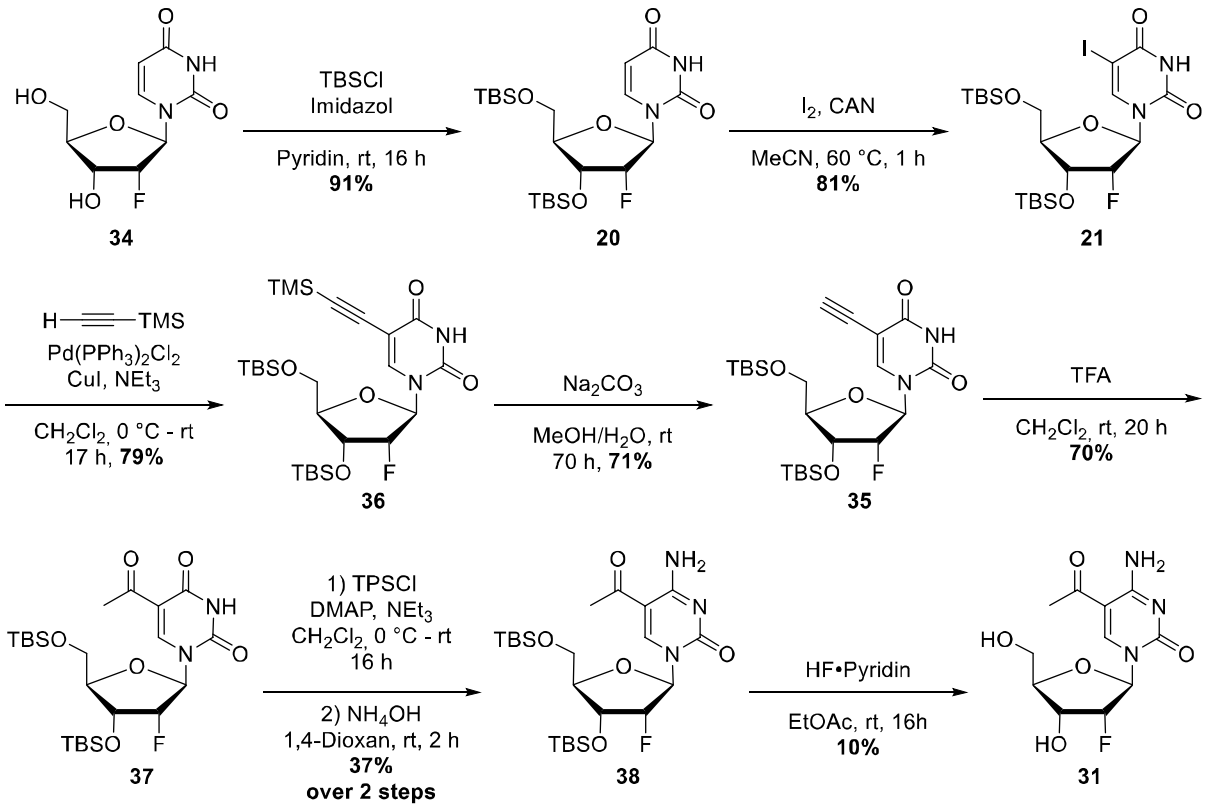
^{13}C-NMR (101 MHz, D_2O/CD_3CN 1:1): δ/ppm = 166.7 (C_a), 166.6 (C4), 155.9 (C2), 150.0 (C6), 97.4 (C5), 94.9 (d, $^1J_{C–F}$ = 185.3 Hz, C2'), 90.1 (d, $^2J_{C–F}$ = 34.1 Hz, C1'), 83.3 (C4'), 67.7 (d, $^2J_{C–F}$ = 16.5 Hz, C3'), 59.3 (C5'), 53.1 (C_b).

^{19}F-NMR (377 MHz, D_2O/CD_3CN 1:1): δ/ppm = −202.5 (ddd, $^{2}J_{F-H}$ = 52.5 Hz, $^{3}J_{F-H}$ = 25.0 Hz, $^{3}J_{F-H}$ = 16.7 Hz C2'-*F*).

HRMS (ESI): berechnet für $C_{11}H_{15}FN_3O_6$ $[M+H]^+$ 304.0939, gefunden: 304.0937.

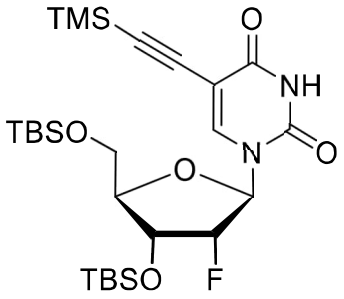
IR (ATR): v/cm^{-1} = 3412 (w), 3103 (w), 2857 (w), 2688 (w), 1710 (s), 1673 (s), 1501 (m), 1471 (w), 1463 (w), 1376 (w), 1361 (w), 1340 (w), 1321 (m), 1280 (w), 1250 (s), 1191 (m), 1114 (s), 1091 (s), 1075 (s), 974 (w), 938 (w), 912 (w), 89 9 (w), 865 (w), 840 (vs), 815 (m), 790 (s), 774 (vs), 723 (w), 662 (m).

5.2.5 Synthese von 5-Acetyl-2'-desoxy-2'-(*R*)-fluoro-cytidin (31)



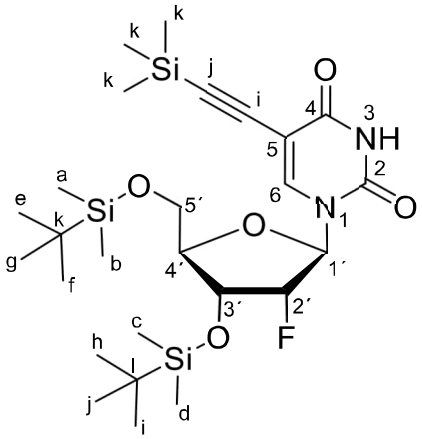
Schema 5.4: Syntheseschema des 2'-F-fdC Analogon 5-COMe-2'-F-dC (**31**) ausgehend von kommerzeill erwärblichem (**34**) 2'F-dU.

5.2.5.1 3',5'-Bis-*O*-(*tert*-butyl(dimethyl)silyl)-5-(2-trimethylsilylethynyl-2'-desoxy-2'-(*R*)-fluoro-udirin (36)



36

Nach einer modifizierten Literaturvorschrift[201] wurde das Nukleosid **21** (1.66 g, 2.77 mmol, 1.10 Äq.) in CH_2Cl_2 (13 mL) gelöst, auf 0 °C abgekühlt und mit Bis(triphenylphosphin)palladium(II) dichlorid (88.4 mg, 126 µmol, 5.00 mol%) und CuI (24.0 mg, 126 µmol, 5.00 mol%) versetzt. Anschließend wurde die Reaktionsmischung mit Triethylamin (2.55 g, 3.5 mL, 25.2 mmol, 10.0 Äq.) und (Trimethylsilyl)acetylen (248 mg, 349 µL, 2.52 mmol, 1.00 Äq.) versetzt und für 17 h bei Raumtemperatur gerührt. Die Lösung wurde mit einer gesättigten, wässrigen $NaHCO_3$-Lösung (12 mL) versetzt und mit CH_2Cl_2 (4 x 100 mL) extrahiert. Die vereinigten organsichen Phasen wurden mit gesättigter, wässriger NaCl-Lösung (50 mL) gewaschen und über Na_2SO_4 getrocknet. Flüchtige Bestandeteile wurden unter vermindertem Druck entfernt und das Rohprodukt wurde säulenchromatographisch (SiO_2, *i*Hexan/EtOAc = 4:1) aufgereinigt. Verbindung **36** (1.25 g, 2.19 mmol, 79%) wurde in Form eines leicht-gelben Schaums erhalten.



$\mathbf{R_f}$ = 0.38 (SiO_2, *i*Hexan/EtOAc = 4:1).

^{1}H-NMR (599 MHz, $CDCl_3$): δ/ppm = 8.01 (s, 1H, N3-H), 7.87 (s, 1H, C6-H), 6.11 (dd, $^{3}J_{H-F}$ = 14.0 Hz, ^{3}J = 3.8 Hz, 1H, C1'-H), 4.81 (dt, $^{2}J_{H-F}$ = 52.6 Hz, ^{3}J = 4.3 Hz, ^{3}J = 4.3 Hz, 1H, C2'-H), 4.29 (dt, ^{3}J= 12.4 Hz, ^{3}J = 4.9 Hz, ^{3}J = 4.9 Hz, 1H, C3'-H), 4.10–4.07 (m, 1H, C4'-H), 3.99 (dd, ^{2}J = 11.8 Hz, ^{3}J = 1.7 Hz, 1H, C5'-H_a)*, 3.77 (dd, ^{2}J = 11.8 Hz, ^{3}J = 1.8 Hz, 1H, C5'-H_b)*, 0.96 (s, 9H, C_{e-g}-H_3)**, 0.91 (s, 9H, C_{i-j}-H_3)**, 0.22 (s, 9H, $C_{k'}$-H_3), 0.17 (s, 3H, C_a-H_3)***, 0.16 (s, 3H, C_b-H_3)***, 0.12 (s, 3H, C_c-H_3)***, 0.11 (s, 3H, C_d-H_3)***.

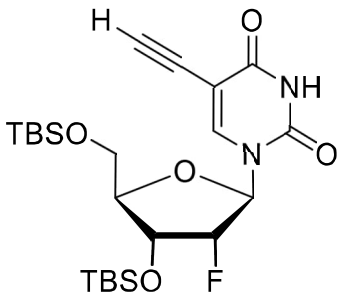
^{13}C-NMR (101 MHz, $CDCl_3$): δ/ppm = 160.7 (C4), 148.9 (C2), 142.1 (C6), 101.0 (C5), 100.5 ($C_{j'}$), 95.0 ($C_{i'}$), 92.6 (d, $^{1}J_{C-F}$ = 195.7 Hz, C2'), 87.4 (d, $^{2}J_{C-F}$ = 33.4 Hz, C1'), 85.3 (d, $^{3}J_{C-F}$ = 1.4 Hz, C4'), 69.7 (d, $^{2}J_{C-F}$ = 15.5 Hz, C3'), 61.7 (C5'), 26.3 (3C, C_{e-g})*, 25.8 (3C, C_{h-j})*, 18.7 (C_k)**, 18.3 (C_l)**, 0.0 ($C_{k'}$), −4.5 (C_a)***, −5.0 (2C, $C_{b,c}$)***, −5.2 (C_d)***.

^{19}F-NMR (377 MHz, $CDCl_3$): δ/ppm = −205.5 (dt, $^{2}J_{F-H}$ = 53.0 Hz, $^{3}J_{F-H}$ = 13.7 Hz, $^{3}J_{F-H}$ = 13.7 Hz, C2'-*F*).

HRMS (ESI): berechnet für $C_{26}H_{48}FN_2O_5Si_3$ $[M+H]^+$ 571.2850, gefunden: 571.2856.

IR (ATR): v/cm^{-1} = 3171 (w), 3063 (w), 2953 (w), 2929 (w), 2902 (w), 2858 (w), 2135 (w), 1723 (s), 1685 (s), 1615 (w), 1471 (w), 1455 (m), 1392 (w), 1360 (w), 1330 (w), 1317 (w), 1276 (m), 1276 (m), 1250 (s), 1232 (w), 1187 (w), 1161 (m), 1128 (s), 1079 (s), 1047 (m), 998 (s), 968 (m), 926 (w), 879 (m), 856 (s), 842 (vs), 777 (vs), 758 (s), 704 (m), 687 (m), 665 (m).

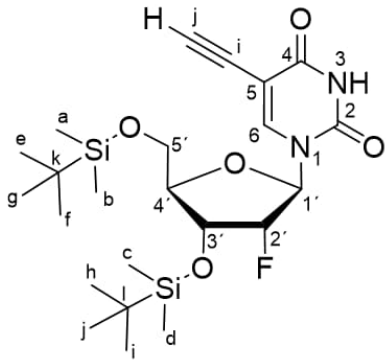
5.2.5.2 3',5'-Bis-*O*-(*tert*-butyl(dimethyl)silyl)-5-ethynyl-2'-desoxy-2'-(*R*)-fluoro-udirin (35)



35

Nukleosid **36** (1.25 g, 2.20 mmol, 1.00 Äq.) wurde in Methanol (44 mL) gelöst und mit einer Lösung von Na_2CO_3 (465 mg, 4.39 mmol, 2.00 Äq.) in Wasser (17 mL) versetzt. Die Reaktionsmischung wurde 69 h bei Raumtemperatur gerührt, anschließend mit gesättigter, wässriger NH_4Cl-Lösung (150 mL) versetzt. Methanol wurde unter vermindertem Druck

entfernt und die wässrige Lösung mit EtOAc (3 x 50 mL) extrahiert. Die vereinigten organischen Phasen wurden mit gesättigter, wässriger NaCl-Lösung (50 mL) gewaschen und über Na_2SO_4 getrocknet. Säulenchromatographische Aufreinigung (SiO_2, *i*Hexan/EtOAc = 5:1) ergab die Verbindung **35** (780 mg, 1.56 mmol, 71%) in Form eines leicht braunen Schaumes.



R_f = 0.17 (SiO_2, *i*Hexan/EtOAc = 5:1).

^{1}H-NMR (599 MHz, $CDCl_3$): δ/ppm = 8.04 (s, 1H, N3-H), 8.03 (s, 1H, C6-H), 6.11 (dd, $^3J_{H–F}$ = 14.3 Hz, ^3J = 3.3 Hz, 1H, C1'-H), 4.80 (dt, $^2J_{H–F}$ = 52.4 Hz, ^3J = 3.6 Hz, ^3J = 3.6 Hz, 1H, C2'-H), 4.29 (dt, $^3J_{H–F}$ = 14.1 Hz, ^3J = 5.4 Hz, ^3J = 5.4 Hz, 1H, C3'-H), 4.10 (dq, ^3J = 5.6 Hz, ^3J = 1.9 Hz, ^3J = 1.9 Hz, 1H, C4'-H), 4.02 (dd, ^2J = 11.8 Hz, ^3J = 1.8 Hz, 1H, C5'-H_a)*, 3.78 (dd, ^2J = 11.8 Hz, ^3J = 1.8 Hz, 1H, C5'-H_b)*, 3.19 (s, 1H, $C_{j'}$-H), 0.95 (s, 9H, C_{e-g}-H_3)**, 0.91 (s, 9H, C_{h-j}-H_3)**, 0.16 (s, 3H, C_a-H_3)***, 0.16 (s, 3H, C_b-H_3)***, 0.13 (s, 3H, C_c-H_3)***, 0.11 (s, 3H, C_d-H_3)***.

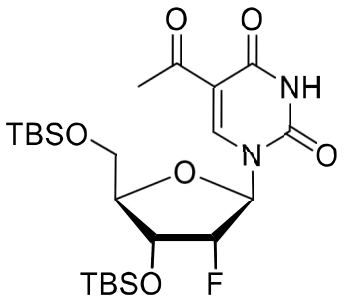
^{13}C-NMR (151 MHz, $CDCl_3$): δ/ppm = 160.8 (C4), 148.8 (C2), 143.0 (C6), 99.7 (C5), 92.8 (d, $^1J_{C–F}$ = 195.3 Hz, C2'), 87.6 (d, $^2J_{C–F}$ = 33.6 Hz, C1'), 85.1 (d, $^3J_{C–F}$ = 1.5 Hz, C4'), 82.9 ($C_{j'}$), 74.6 ($C_{i'}$), 69.5 (d, $^2J_{C–F}$ = 15.7 Hz, C3'), 61.5 (C5'), 26.3 (3C, C_{e-g})*, 25.8 (3C, C_{h-j})*, 18.8 (C_k)**, 18.3 (C_l)**, −4.5 (d, C_a)***, −4.9 (C_b)***, −5.1 (C_c)***, −5.2 (C_d)***.

^{19}F-NMR (377 MHz, $CDCl_3$): δ/ppm = −204.7 (dt, $^2J_{F–H}$ = 52.8 Hz, $^3J_{F–H}$ = 14.2 Hz, $^3J_{F–H}$ = 14.2 Hz, C2'-*F*).

HRMS (ESI): berechnet für $C_{23}H_{40}FN_2O_5Si_2$ $[M+H]^+$ 499.2454, gefunden: 499.2455.

IR (ATR): v/cm^{-1} = 3316 (bw), 3195 (bw), 3086 (w), 2954 (w), 2797 (w), 2904 (w), 2857 (w), 1698 (s), 1622 (w), 1471 (w), 1453 (m), 1408 (w), 1391 (w), 1361 (w), 1254 (s), 1224 (m), 1156 (m), 1120 (s), 1074 (s), 996 (m), 979 (m), 938 (w), 847 (m), 852 (m), 843 (vs), 778 (vs), 702 (w), 670 (m).

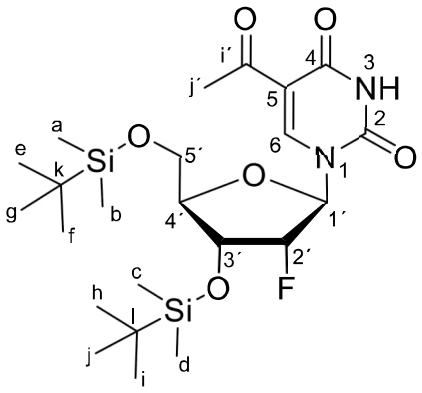
5.2.5.3 3',5'-Bis-*O*-(*tert*-butyldimethylsilyl)-5-acetyl-2'-desoxy-2'-(*R*)-fluoro-uridin (37)



37

Nukleosid **35** (390 mg, 782 µmol, 1.00 Äq.) wurde in CH_2Cl_2 (44 mL) gelöst und tropfenweise (0.10 mL/30 min) mit Trifluoressigsäure (1.16 g, 784 µL, 10.2 mmol, 13.0 Äq.) versetzt. Reaktionsmischung wurde für 18 h bei Raumtemperatur gerührt, anschließend wurde die Lösung mit gesättigter, wässriger $NaHCO_3$-Lösung (100 mL) neutralisert. Die wässrige Phase wurde mit CH_2Cl_2 (4 x 100 mL) extrahiert, die vereinigten organischen Phasen wurden mit gesättigter wässriger NaCl-Lösung (100 mL) gewaschen und über Na_2SO_4 getrocknet. Nach dem alle flüchtigen Bestandteile unter vermindertem Druck entfernt wurde, wurde das Rohprodukt säulenchromatographisch (SiO_2, *i*Hexan/EtOAc = 2:1) aufgereinigt. Die analytischen Daten ergaben, dass es sich bei der isolierten Verbindung um ein einfach TBS-entschuützes Nukleosid handelt. Im Weiteren wurde aus diesem Grund eine Schützung der freien OH Gruppe wie folgt durchgeführt.

Das erhaltene teilentschützte Nukleosid (221 mg, 548 µmol, 1.00 Äq.) wurde in Pyridin (3.5 mL) gelöst, mit Imidazol (138.0 mg, 2.22 µmol, 3.70 Äq.) und TBS-Cl (202 mg, 1.36 mmol, 2.50 Äq.) versetzt und für 18 h bei Raumtemperatur gerührt. Nach vollständigem Umsatz wurde die eine gesättigte, wässrige $NaHCO_3$-Lösung (50 mL) zugegeben und Pyridin unter vermindertem Druck entfernt. Der Rückstand wurde in Wasser (50 mL) aufgenommen und mit EtOAc (4 x 100 mL) extrahiert. Die vereinigten organischen Phasen wurden mit gesättigter wässriger NaCl-Lösung (100 mL) gewaschen und über Na_2SO_4 getrocknet. Säulenchrimatographische Aufreinigung (SiO_2, *i*Hexan/EtOAc = 4:1) ergab die Verbindung **37** (171 mg, 330 µmol, 60%) als farbloßen Schaum.



R_f = 0.54 (SiO_2, *i*Hexan/EtOAc = 2:1).

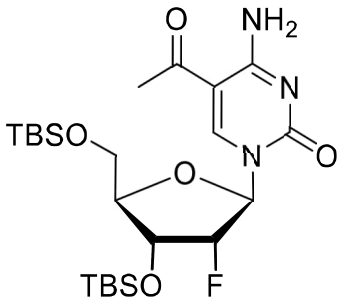
^{1}H-NMR (800 MHz, $CDCl_3$): δ/ppm = 8.41 (s, 1H, C6-H), 7.98 (s, 1H, N3-H), 6.07 (dd, $^3J_{H–F}$ = 14.5 Hz, ^3J = 4.2 Hz, 1H, C1'-H), 4.92 (dt, $^2J_{H–F}$ = 52.8 Hz, ^3J = 4.5 Hz, ^3J = 4.5 Hz, 1H, C2'-H), 4.34 (dt, $^3J_{H–F}$ = 10.5 Hz, ^3J = 4.8 Hz, ^3J = 4.8 Hz, 1H, C3'-H), 4.11–4.09 (m, 1H, C4'-H), 3.94 (dd, ^2J = 11.7 Hz, ^3J = 2.2 Hz, 1H, C5'-H_a)*, 3.77 (dd, ^2J = 11.8 Hz, ^3J = 2.9 Hz, 1H, C5'-H_b)*, 2.59 (s, 3H, $C_{j'}$-H_3), 0.92 (s, 9H, C_{e-g}-H_3)**, 0.91 (s, 9H, C_{h-j}-H_3)**, 0.13 (s, 3H, C_a-H_3)***, 0.13 (s, 6H, $C_{b,c}$-H_3)***, 0.11 (s, 3H, C_d-H_3)***.

^{13}C-NMR (101 MHz, $CDCl_3$): δ/ppm = 193.4 ($C_{i'}$), 160.3 (C4), 149.2 (C2), 146.9 (C6), 113.5 (C5), 92.3 (d, $^1J_{C–F}$ = 195.8 Hz, C2'), 88.6 (d, $^2J_{C–F}$ = 33.4 Hz, C1'), 85.8 (d, $^3J_{C–F}$ = 1.3 Hz, C4'), 70.2 (d, $^2J_{C–F}$ = 15.4 Hz, C3'), 62.1 (C5'), 30.8 ($C_{j'}$), 26.2 (3C, C_{e-g})*, 25.8 (3C, C_{h-j})*, 18.7 (C_k)**, 18.3 (C_l)**, −4.6 (d, C_a)***, −5.0 (C_b)***, −5.3 (C_c)***, −5.5 (C_d)***.

HRMS (ESI): berechnet für $C_{23}H_{42}FN_2O_6Si_2$ $[M+H]^+$ 517.2560, gefunden: 517.2566.

IR (ATR): v/cm^{-1} = 3351 (bw), 3192 (bw), 3069 (w), 2954 (w), 2929 (w), 2903 (w), 2857 (w), 1697 (s), 1622 (m), 1453 (m), 1408 (w), 1391 (w), 1361 (w), 1254 (m), 1244 (w), 1156 (m), 1120 (m), 1074 (m), 995 (w), 979 (w), 574 (m), 833 (vs), 778 (vs), 702 (w), 670 (m).

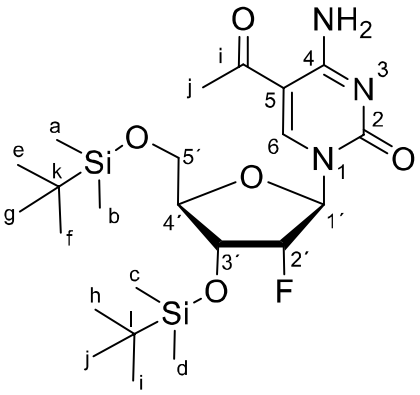
5.2.5.4 3',5'-Bis-*O*-(*tert*-butyldimethylsilyl)-5-acetyl-2'-desoxy-2'-(*R*)-fluoro-cytidin (38)



38

Verbidnung **37** (171 mg, 330 µmol, 1.00 Äq.) wurde in CH_2Cl_2 (3.3 mL) gelöst, auf 0 °C gekühlt und und mit DMAP (8.0 mg, 66.1 µmol, 0.20 Äq.) und NEt_3 (334 mg, 460 µL, 3.30 mmol, 10.0 Äq.) versetzt. Bei 0 °C wurde TPS-Cl (150 mg, 495 µmol, 1.50 eq.) portionsweise über einen Zeitraum von 30 min zugegeben. Die Reaktionsmischung wurde auf Raumtemperatur erwärmen lassen und für 16 h bei Raumtemperatur gerührt. Die Reaktion wurde durch Zugabe einer gesättigten wässrigen NH_4Cl-Lösung (10 mL) neutralisiert, und mit CH_2Cl_2 (4 x 50 mL) extrahiert. Die vereinigten organischen Phasen wurden über Na_2SO_4 getrocknet und die Lösungsmittel wurden *in vacuo* entfernt. Das erhaltene Rohprodukt wurde ohne weitere Aufreinigungsschritte weiter umgesetzt.

Das erhaltene Rohprodukt wurde in 1,4-Dioxan (6.6 mL) gelöst und tropfenweise mit einer wässrigen Lösung von NH_4OH (25% aq., 6.30 mL/mmol, 2.00 mL) versetzt und für 16 h bei Raumtemperatur gerührt. Die Reaktion wurde durch Zugabe einer gesättigten wässrigen NH_4Cl-Lösung (10 mL) beendet und mit EtOAc (3 x 100 mL) extrahiert. Die vereinigten organischen Phasen wurden mit einer gesättigten wässrigen NaCl-Lösung (100 mL) gewaschen und anschließend über Na_2SO_4 getrocknet. Lösungsmittel wurden *in vacuo* entfernt und das Rohprudukt säulenchromatographisch aufgereinigt (SiO_2, CH_2Cl_2/MeOH = 50:1) um die Verbindung **38** in Form eines roten Pulvers (63.0 mg, 122 µmol, 37%) zu erhalten.



R_f = 0.41 (SiO_2, CH_2Cl_2/MeOH = 20:1).

^{1}H-NMR (800 MHz, $CDCl_3$): δ/ppm = 8.73 (s, 1H, C4-NH_2), 8.49 (s, 1H, C6-H), 5.91 (dd, $^3J_{H–F}$ = 17.4 Hz, ^3J = 1.7 Hz, 1H, C1'-H), 5.04 (ddd, $^2J_{H–F}$ = 52.3 Hz, ^3J = 4.5 Hz, ^3J = 1.7 Hz, 1H, C2'-H), 4.22 (ddd, $^3J_{H–F}$ = 18.5 Hz, ^3J = 7.6 Hz, ^3J = 4.6 Hz, 1H, C3'-H), 4.17–4.13 (m, 1H, C4'-H), 4.08 (dd, ^2J = 11.9 Hz, ^3J = 1.8 Hz, 1H, C5'-H_a)*, 3.83 (dd, ^2J = 11.9 Hz, ^3J = 3.8 Hz, 1H, C5'-H_b)*, 2.43 (s, 3H, $C_{j'}$-H_3), 0.91 (s, 9H, C_{e-g}-H_3)**, 0.90 (s, 9H, C_{h-j}-H_3)**, 0.13 (s, 3H, C_a-H_3)***, 0.11 (s, 3H, C_b-H_3)***, 0.10 (s, 3H, C_c-H_3)***, 0.09 (s, 3H, C_d-H_3)***.

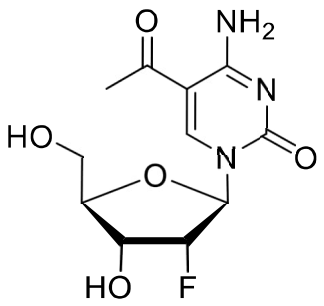
^{13}C-NMR (201 MHz, $CDCl_3$): δ/ppm = 195.2 ($C_{i'}$), 163.6 (C4), 153.4 (C2), 149.5 (C6), 104.6 (C5), 92.7 (d, $^1J_{C–F}$ = 192.3 Hz, C2'), 91.0 (d, $^2J_{C–F}$ = 34.8 Hz, C1'), 84.6 (C4'), 69.4 (d, $^2J_{C–F}$ = 16.7 Hz, C3'), 62.1 (C5'), 26.4 ($C_{j'}$), 26.3 (3C, C_{e-g})*, 25.8 (3C, C_{h-j})*, 18.8 (C_k)**, 18.2 (C_l)**, −4.3 (d, C_a)***, −4.8 (C_b)***, −4.9 (C_c)***, −5.1 (C_d)***.

^{19}F-NMR (377 MHz, $CDCl_3$): δ/ppm = −200.0 (dt, $^2J_{F–H}$ = 52.3 Hz, $^3J_{F–H}$ = 18.3 Hz, $^3J_{F–H}$ = 16.8 Hz, C2'-*F*).

HRMS (ESI): berechnet für $C_{23}H_{43}FN_3O_5Si_2$ $[M+H]^+$ 516.2720, berechnet: 516.2721.

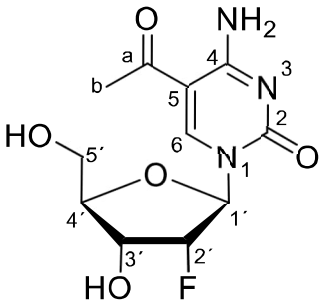
IR (ATR): v/cm^{-1} = 3351 (w), 3192 (w), 3069 (w), 2954 (w), 2929 (w), 2903 (w), 2857 (w), 1697 (s), 1622 (m), 1453 (m), 1408 (w), 1391 (w), 1361 (w), 1254 (m), 1244 (w), 1156 (m), 1120 (m), 1074 (m), 995 (w), 979 (w), 574 (m), 833 (vs), 778 (vs), 702 (w), 670 (m).

5.2.5.5 5-Acetyl-2'-desoxy-2'-(*R*)-fluoro-cytidin (31)



31

In einem Polypropylen Rohr wurde nach der *ALLGEMEINEN REAKTIONSVORSCHRIFT I* Verbindung **38** (63.0 mg, 122 µmol, 1.00 Äq.) in EtOAc (1.83 mL) gelöst und mit HF • Pyridin (70%, 47.6 µL, 1.83 mmol, 15.0 Äq.) versetzt. Die Reaktionsmischung wurde 18 h bei Raumtemperatur gerührt, anschließend auf 0 °C gekühlt, mit TMS • OMe (408 mg, 538 µL, 3.91 mmol, 32.0 Äq.) versetzt und für 1 h bei 0 °C gerührt. Flüchtigen Bestandteile wurden *in vacuo* entfernet und das erhaltene Öl in H_2O (30 mL) aufgenommen. und mit CH_2Cl_2 (3 x 100 mL) extrahiert. Die wässrige Phase wurde unter vermindertem Druck konzentriert und mittels semipräparativer HPL-Chromatographie (MeCN/H_2O, MeCN: 0 → 20%, 45 min, Retentionszeit: 25 min) aufgereinigt. Verbindung **31** (2.68 mg, 9.33 µmol, 10%). wurde nach Lyophilisation als farbloser Feststoff erhalten.



Retentionszeit = 25 min (C_{18}-SiO_2, MeCN/H_2O, MeCN: 0 → 20%, 45 min).

^{1}H-NMR (800 MHz, D_2O): δ/ppm = 9.07 (s, 1H, C6-H), 6.07 (d, $^3J_{H-F}$ = 17.5 Hz, 1H, C1'-H), 5.14 (dd, $^2J_{H-F}$ = 52.4 Hz, ^3J = 4.1 Hz, 1H, C2'-H), 4.39 (ddd, $^3J_{H-F}$ = 24.9 Hz, ^3J = 4.1 Hz, ^3J = 9.4 Hz, 1H, C3'-H), 4.22 (dt, ^3J = 9.4 Hz, ^3J= 2.2 Hz, 1H, C4'-H), 4.12 (dd, ^2J = 13.1 Hz, ^3J = 2.3 Hz, 1H, C5'-H_a)*, 3.89 (dd, ^2J = 13.1 Hz, ^3J = 2.5 Hz, 1H, C5'-H_b)*, 2.46 (s, 3H, C_b-H_3).

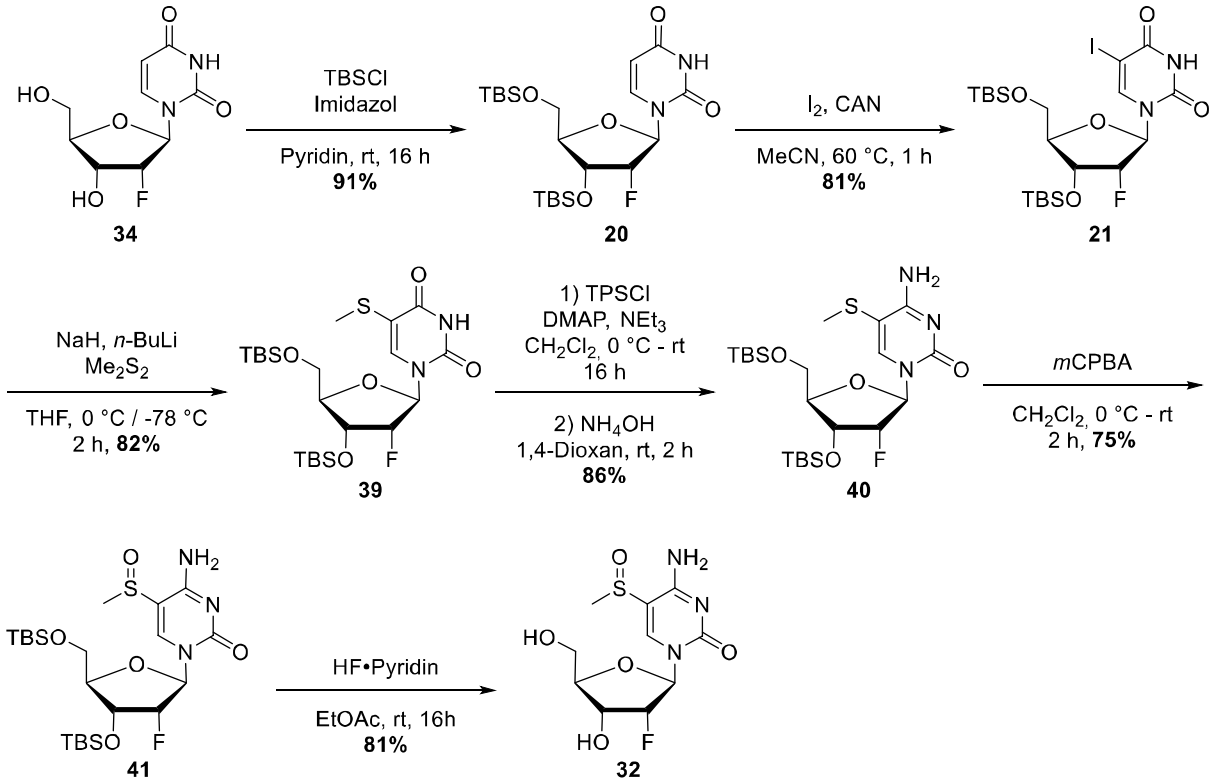
^{13}C-NMR (201 MHz, D_2O): δ/ppm = 198.9 (C_a), 163.6 (C4), 155.1 (C2), 150.6 (C6), 104.9 (C5), 93.8 (d, $^1J_{C2'–F2'}$ = 185.2 Hz, C2'), 89.6 (d, $^2J_{C1'–F2'}$ = 34.7 Hz, C1'), 82.1 (C4'), 66.8 (d, $^2J_{C3'–F2'}$ = 16.6 Hz, C3'), 58.3 (C5'), 25.8 (C_b).

^{19}F-NMR (377 MHz, D_2O): δ/ppm = −201.6 (ddd, $^2J_{F–H}$ = 52.9 Hz, $^3J_{F–H}$ = 17.7 Hz, $^3J_{F–H}$ = 24.9 Hz, C2'-*F*).

HRMS (ESI): berechnet für $C_{11}H_{15}FN_3O_5$ $[M+H]^+$ 288.0990, gefunden: 288.0990.

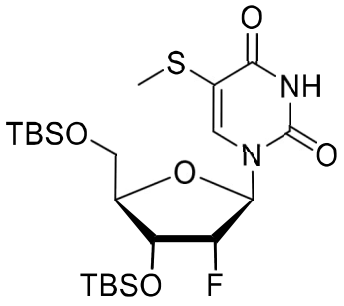
IR (ATR): v/cm^{-1} = 3334 (bw), 3207 (bw), 2928 (w), 2227 (w), 1697 (s), 1644 (s), 1491 (m), 1408 (w), 1295 (w), 1252 (w), 1096 (s),1059 (m), 984 (w), 949 (w), 845 (m), 781 (m), 705 (w).

5.2.6 Synthese von 5-(Methylsulfinyl)-2'-desoxy-2'-(*R*)-fluoro-cytidin (32)



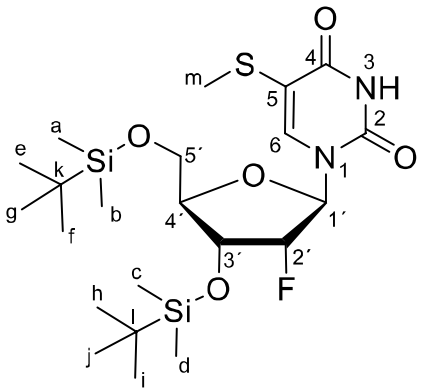
Schema 5.5: Syntheseschema des 2'-F-fdC Analogon 5-SOMe-2'-F-dC (**32**) ausgehend von kommerziell erwärblichem 2'F-dU (**34**).

5.2.6.1 3',5'-Bis-*O*-(*tert*-butyldimethylsilyl)-5-(methylthio)-2'-desoxy-2'-(*R*)-fluoro-uridin (39)



39

Analog zu einer modifizierten Literaturvorschrift[202] wurde das 5-iodierte Uridinderivat **21** (0.20 g, 0.33 mmol, 1.0 Äq.) in THF (2.0 mL) gelöst, auf 0 °C abgekühlt und mit NaH (20 mg, 0.50 mmol, 1.5 Äq.) versetzt. Die Reaktionslösung wurde 30 min bei Raumtemeratur gerührt. Anschließend wurde die Reaktion auf -78 °C gekühlt und *n*-BuLi (2.5 M, 0.40 mL, 1.0 mmol, 3.0 Äq.) wurde tropfenweise zugegeben. Die resuliterende gelbe Lösung wurde 15 min bei -78 °C gerührt und mit Me_2S_2 (0.18 mL, 2.0 mmol, 6.0 Äq.) versetzt. Die Reaktion wurde weitere 60 min bei -78 °C gerührt, anschließend mit H_2O (10 mL) versetzt und CH_2Cl_2 (10 mL) verdünnt. Die wässrige Phase wurde mit CH_2Cl_2 (4 x 20 mL) extrahiert, die vereinigten organischen Phasen mit gesättigter, wässriger NaCl-Lösung (2 x 100 mL) gewaschen und über Na_2SO_4 getrocknet. Flüchtige Bestandteile wurden unter vermindertem Druck entfernt und das Rohprodukt wurde säulenchromatographisch (SiO_2, iHexan/EtOAc = 5:1) aufgereinigt um die Verbindung **39** (0.15 g, 0.27 mmol. 82 %) in Form eines farblosen Feststoffes zu erhalten.



R_f = 0.15 (SiO_2, iHexan/EtOAc = 5:1).

^{1}H-NMR (599 MHz, $CDCl_3$): δ/ppm = 8.12 (s, 1H, N3-H), 7.91 (s, 1H, C6-H), 6.08 (dd, $^{3}J_{H\text{-}F}$ = 14.7 Hz, ^{3}J = 3.8 Hz, 1H, C1'-H), 4.84 (dt, $^{2}J_{H\text{-}F}$ = 52.7 Hz, ^{3}J = 4.2 Hz, 1H, C2'-H), 4.31 (dt, $^{3}J_{H\text{-}F}$ = 12.4 Hz, 3J = 5.0 Hz, 1H, C3'-H), 4.07 (dq, ^{3}J = 5.4 Hz, ^{3}J = 2.0 Hz, 1H, C4'-H), 3.99 (dd, ^{2}J = 11.8 Hz, ^{3}J = 2.0 Hz, 1H, C5'-H_a)*, 3.77 (dd, ^{3}J = 11.8 Hz, ^{3}J = 2.3 Hz, 1H, C5'-H_b)*, 2.32 (s, 3H, C_m-H_3), 0.96 (s, 9H, $C_{e\text{-}g}$-H_3)**, 0.92 (s, 9H, $C_{h\text{-}j}$-H_3)**, 0.17 (s, 3H, C_a-H_3)***, 0.16 (s, 3H, C_b-H_3)***, 0.13 (s, 3H, C_c-H_3)***, 0.11 (s, 3H, C_d-H_3)***.

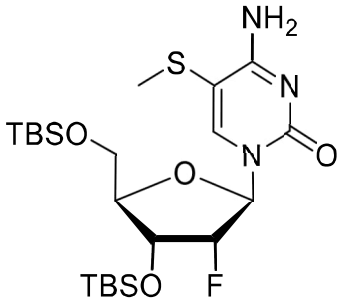
^{13}C-NMR (101 MHz, $CDCl_3$): δ/ppm = 161.1 (C4), 149.8 (C2), 141.8 (C6), 110.6 (C5), 92.5 (d, $^{1}J_{C\text{–}F}$ = 194.9 Hz, C2'), 87.6 (d, $^{2}J_{C\text{–}F}$ = 33.4 Hz, C1'), 85.2 (C4'), 69.8 (d, $^{2}J_{C\text{–}F}$ = 15.4 Hz, C3'), 61.9 (C5'), 26.3 (3C, $C_{e\text{-}g}$)*, 25.8 (3C, $C_{h\text{-}j}$)*, 18.8 (C_k)**, 18.3 (C_l)**, 17.8 (C_m), −4.5 (d, C_a)***, −4.9 (C_b)***, −5.1 (C_c)***, −5.1 (C_d)***.

^{19}F-NMR (377 MHz, $CDCl_3$): δ/ppm = −205.1 (dt, $^{2}J_{F\text{–}H}$ = 52.9 Hz, $^{3}J_{F\text{–}H}$ = 13.6 Hz, C2'-*F*).

HRMS (ESI): berechnet für $C_{22}H_{42}FN_2O_5SSi_2$ $[M+H]^+$ 521.2332, gefunden: 521.2331.

IR (ATR): v/cm^{-1} = 3315 (w), 3069 (w), 2954 (w); 2929 (w), 2903 (w), 2857 (w), 1697 (s), 1622 (w), 1471 (w), 1453 (m), 1408 (w), 1361 (w), 1254 (m), 1156 (m), 1120 (s), 1074 (s), 995 (w), 979 (w), 874 (m), 857 (m), 833 (vs), 778 (vs), 702 (w), 670 (m).

5.2.6.2 3',5'-Bis-*O*-(*tert*-butyldimethylsilyl)-5-(methylthio)-2'-desoxy-2'-(*R*)-fluoro-cytidin (40)

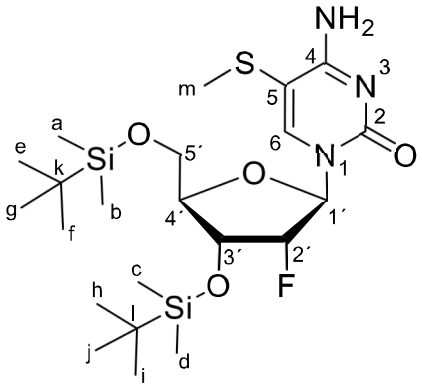


40

Verbindung **39** (0.10 g, 0.19 mmol, 1.0Äq.) wurde in CH_2Cl_2 (2.0 mL) gelöst, auf 0 °C gekühlt und und mit DMAP (5.0 mg, 40 µmol, 0.2 Äq.) und NEt_3 (0.19 g, 260 µL, 1.9 mmol, 10.0 Äq.) versetzt. Bei 0 °C wurde TPS-Cl (85 mg, 0.28 mmol, 1.5 eq.) portionsweise über einen Zeitraum von 30 min zugegeben. Die Reaktionsmischung wurde auf Raumtemperatur erwärmen lassen und für 16 h bei Raumtemperatur gerührt. (Umsatz wurde mittels DC-

Kontrolle bestimmt und mit Hilfe von einer *n*-Butylamin-Lösung als nukleophiles Abfangreagenz zur besseren Sichtbarkeit auf der DC-Platte, durchgeführt. *i*Hexan/EtOAc = 2:1, n-Butylaminaddukt R_f = 0.5) Die Reaktion wurde mit CH_2Cl_2 (20 mL) verdünnt und durch Zugabe von H_2O (20 mL) abgebrochen. Die wässrige Lösung wurde mit CH_2Cl_2 (3 x 30 mL) extrahiert. Die vereinigten organischen Phasen wurden mit einer gesättigten, wässrigen NaCl-Lösung (3 x 90 mL) gewaschen und über Na_2SO_4 getrocknet und die Lösungsmittel wurden *in vacuo* entfernt.

Das erhaltene Rohprodukt wurde in 1,4-Dioxan (3.7 mL) gelöst und tropfenweise mit einer wässrigen Lösung von NH_4OH (25% aq., 6.3 mL/mmol, 1.2 mL) versetzt und für 19 h bei Raumtemperatur gerührt. Die Reaktion wurde durch Zugabe einer gesättigten wässrigen NH_4Cl-Lösung (20 mL) beendet und mit EtOAc (4 x 30 mL) extrahiert. Die vereinigten organischen Phasen wurden mit einer gesättigten wässrigen NaCl-Lösung (2 x 100 mL) gewaschen und anschließend über Na_2SO_4 getrocknet. Lösungsmittel wurden *in vacuo* entfernt und das Rohprudukt säulenchromatographisch aufgereinigt (SiO_2, CH_2Cl_2/MeOH = 40:1) um die Verbindung **40** in Form eines gelben Feststoffes (83.0 mg, 160 µmol, 86%) zu erhalten.



R_f = 0.31 (SiO_2, CH_2Cl_2/MeOH = 40:1).

^{1}H-NMR (800 MHz, $CDCl_3$): δ/ppm = 8.67 (s, 1H, C4-NH_2), 8.08 (s, 1H, C6-H), 6.00 (dd, $^3J_{H–F}$ = 16.8 Hz, ^3J = 1.6 Hz, 1H, C1'-H), 4.82 (ddd, $^2J_{H–F}$ = 52.5 Hz, ^3J = 4.3 Hz, ^3J = 1.6 Hz, 1H, C2'-H), 4.22 (ddd, $^3J_{H–F}$ = 20.0 Hz, ^3J = 7.9 Hz, ^3J = 4.3 Hz, 1H, C3'-H), 4.11–4.04 (m, 2H, C4'-H, C5'-H_a*), 3.80 (dd, ^2J = 12.1 Hz, ^3J = 2.5 Hz, 1H, C5'-H_b)*, 2.19 (s, 3H, C_m-H_3), 0.96 (s, 9H, C_{e-g}-H_3)**, 0.89 (s, 9H, C_{h-j}-H_3)**, 0.17 (s, 3H, C_a-H_3)***, 0.16 (s, 3H, C_b-H_3)***, 0.10 (s, 3H, C_c-H_3)***, 0.08 (s, 3H, C_d-H_3)***.

^{13}C-NMR (201 MHz, $CDCl_3$): δ/ppm = 166.1 (C4), 155.2 (C2), 146.4 (C6), 100.6 (C5), 93.3 (d, $^1J_{C–F}$ = 192.5 Hz, C2'), 89.2 (d, $^2J_{C–F}$ = 33.8 Hz, C1'), 83.6 (C4'), 68.6 (d, $^2J_{C–F}$ = 16.3 Hz, C3'),

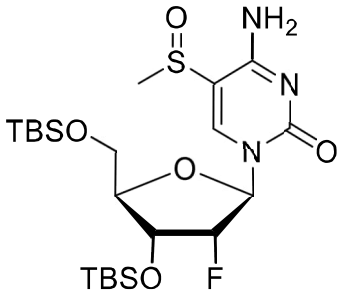
61.0 (C5'), 26.4 (3C, C_{e-g})*, 25.8 (3C, C_{h-j})*, 19.8 (C_m), 18.9 (C_k)**, 18.2 (C_l)**, −4.5 (d, C_a)***, −4.9 (C_b)***, −5.0 (C_c)***, −5.2 (C_d)***.

^{19}F-NMR (377 MHz, $CDCl_3$): δ/ppm = -−201.4 (ddd, $^2J_{F-H}$ = 52.5 Hz, $^3J_{F-H}$ = 20.0 Hz, $^3J_{F-H}$ = 16.7 Hz, C2'-*F*).

HRMS (ESI): berechnet für $C_{22}H_{43}FN_3O_5SSi_2$ $[M+H]^+$ 520.2491, berechnet: 520.2487.

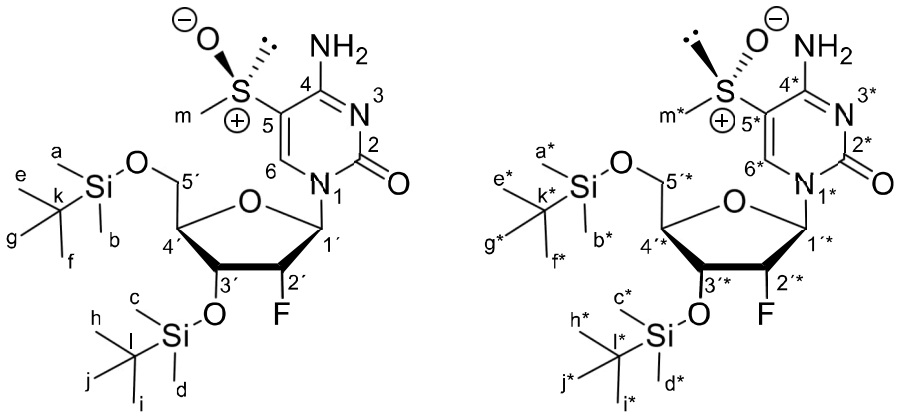
IR (ATR): v/cm^{-1} = 3171 (w), 3062 (w), 2954 (w); 2928 (w), 2887 (w), 2857 (w), 1690 (s), 16042 (w), 1471 (w), 1462 (m), 1440 (w), 1390 (w), 1361 (w), 1322 (w), 1253 (s), 1225 (w), 1165 (m), 1133 (s), 1076 (s), 1054 (m), 996 (w), 979 (w), 938 (w), 893 (w), 858 (m), 830 (vs), 777 (vs), 707 (w), 670 (m).

5.2.6.3 3',5'-Bis-*O*-(*tert*-butyldimethylsilyl)-5-(methylsulfinyl)-2'-desoxy-2'-(*R*)-fluoro-cytidin (41)



41

5-Methylthiocytidin **40** (0.17 g, 0.33 mmol, 1.0 Äq.) wurde in CH_2Cl_2 (1.7 mL) gelöst, auf 0 °C abgekühlt und mit *m*-CPBA (77%, 70 mg, 0.33 mmol, 1.0 Äq.) in mehreren Portionen versetz. Die Reaktion wurde für 1 h bei 0 °C und 2 h bei Raumtemperatur gerührt. Die Reaktion wurde durch Zugabe einer gesättigten wässrigen $NaHCO_3$-Lösung (10 mL) gestoppt und mir CH_2Cl_2 (3 x 20 L) extrahiert. Die vereinigten organischen Phasen wurden mit einer gesättigten, wässrigen NaCl-Lösung (3 x 60 mL) gewaschen und über Na_2SO_4 getrocknet. Säulenchromatigraphsiche Aufreinigung (SiO_2, EtOAc/*i*Hexan/NEt_3 = 5:1:2%) ergab die Verbindung **41** als ein Enantiomerengemisch in Form eines farblosen Feststoffes (0.13 g, 0.25 mmol, 75%).



R_f = 0.18 (SiO_2, EtOAc/*i*Hexan/NEt_3 = 5:1:2%).

^{1}H-NMR (599 MHz, $CDCl_3$): δ/ppm = 8.19 (s, 1H, C6-H), 8.12 (s, 1H, C6*-H), 7.63 (s, 1H, N3-H), 7.56 (s, 1H, N3*-H) 6.00 (dd, $^3J_{H-F}$ = 16.0 Hz, ^3J = 1.0 Hz, 1H, C1'-H), 5.97 (dd, $^3J_{H-F}$ = 16.5 Hz, ^3J = 1.5 Hz, 1H, C1'*-H), 4.85 (dd, $^2J_{H-F}$ = 51.8 Hz, ^3J = 4.3 Hz, 1H, C2'-H), 4.84 (dd, $^2J_{H-F}$ =52.3 Hz, ^3J = 4.0 Hz, 1H, C2'*-H), 4.21-4.07 (m, 6H, C3'-H, C3'*-H, C4'-H, C4'*-H, C5'-H_a*, C5'*-H_a*), 3.80 (dd, ^2J = 12.2 Hz, ^3J = 5.5 Hz, 1H, C5'-H_b)*, 3.80 (dd, ^2J = 12.2 Hz, ^3J = 5.5 Hz, 1H, C5'*-H_b)* 2.93 (s, 3H, C_m-H_3), 2.93 (s, 3H, C_m*-H_3), 0.95 (s, 9H, C_{e-g}-H_3)**, 0.95 (s, 9H, C_{e-g}*-H_3)**, 0.89 (s, 9H, C_{h-j}-H_3)**, 0.89 (s, 9H, C_{h-j}*-H_3)**, 0.15 (s, 3H, C_a-H_3)***, 0.15 (s, 3H, C_a*-H_3)***, 0.15 (s, 3H, C_b-H_3)***, 0.12 (s, 3H, C_b*-H_3)***, 0.11 (s, 3H, C_c-H_3)***, 0.10 (s, 3H, C_c*-H_3)***, 0.09 (s, 3H, C_d-H_3)***, 0.07 (s, 3H, C_d*-H_3)***.

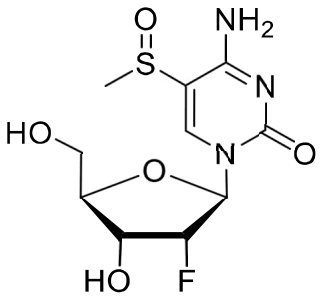
^{13}C-NMR (151 MHz, $CDCl_3$): δ/ppm = 163.1 (C4), 163.0 (C4*), 153.7 (C2), 153.7 (C2*), 141.4 (C6), 141.1 (C6*), 106.1 (C5), 105.7 (C5*), 93.1 (d, $^1J_{C-F}$ = 192.5 Hz, C2'), 92.9 (d, $^1J_{C-F}$ = 193.0 Hz, C2'*), 89.4 (d, $^2J_{C-F}$ = 34.4 Hz, C1'), 89.0 (d, $^2J_{C-F}$ = 34.4 Hz, C1'*), 83.6 (C4'), 83.5 (C4'*), 68.2 (d, $^2J_{C-F}$ = 16.4 Hz, C3'), 67.8 (d, $^2J_{C-F}$ = 16.6 Hz, C3'*), 60.8 (C5'), 60.6 (C5'*), 40.3 (C_m), 40.3 (C_m*), 26.4 (6C, C_{e-g}, C^*_{e-g}), 25.8 (3C, C_{h-j}), 25.8 (3C, C^*_{h-j}), 19.0 (2C, C_k, C^*_k)*, 18.2 (2C, C_l, C^*_l)*, −4.3 (C_a)**, −4.4 (C^*_a)**, −4.7 (C_b)**, −4.8 (C^*_b)**. −4.8 (C_c)**, −4.8 (C^*_c)**, −5.1 (C_d)**, −5.2 (C^*_d)**.

^{19}F-NMR (377 MHz, $CDCl_3$): δ/ppm = −198.4, −206.2 (m, C2'-*F*).

LRMS (ESI): berechnet für $C_{22}H_{43}FN_3O_5SSi_2$ $[M+H]^+$ 536.2440, gefunden: 536.3 berechnet für $C_{22}H_{41}FN_2O_5SSi_2$ $[M-H]^-$ 534.2295, berechnet: 534.3

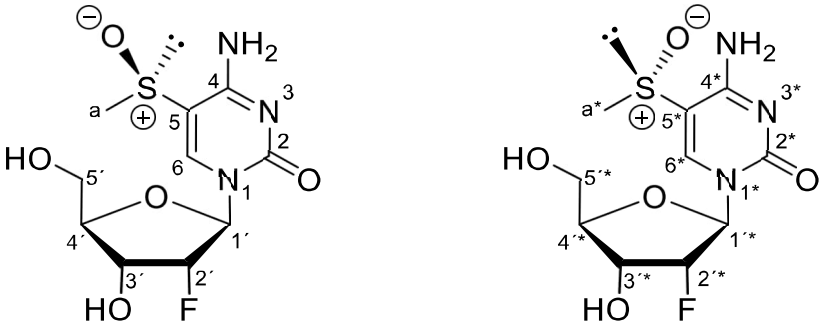
IR (ATR): v/cm^{-1} = 3171 8BW9; 3062 8BW9; 2954 (w), 2928 (w), 2893 (w), 2857 (w), 1691 (s), 1604 (w), 1439 (m), 1225 (w), 1165 (m), 1133 (m), 119 (m), 1077 (m), 1054 (m), 996 (w), 678 (w), 829 (w), 858 (m), 858 (m), 830 (vs), 776 (vs), 706 (w), 670 (m).

5.2.6.4 5-(Methylsulfinyl)-2'-desoxy-2'-(*R*)-fluoro-cytidin (32)



32

Nach der *ALLGEMEINEN REAKTIONSVORSCHRIFT I* wurde **41** (0.11 g, 0.21 mmol, 1.00 Äq.) in EtOAc (3.1 mL) mit HF • Pyridin (70%, 90.0 mg, 82.0 µL, 3.10 mmol, 15.0 Äq.) versetzt. Die Reaktionsmischung wurde 18 h bei Raumtemperatur gerührt, anschließend auf 0 °C gekühlt, mit TMS • OMe (0.70 g, 0.92 mL, 6.70 mmol, 32.0 Äq.) versetzt und für 1 h bei 0 °C gerührt. Aufreinigung mittels semipräparativer HPL-Chromatographie (MeCN/H_2O, MeCN: 0 → 15%, 45 min, Retentionszeit: 20 min, Enantiomerengemisch) ergab Verbindung **21** (52.0 mg, 0.17 mmol, 81%) als Enantiomerengemisch in Form eines farblosen Pulvers.



Retentionszeit = 20 min (C_{18}-SiO_2, MeCN/H_2O, MeCN: 0 → 15%, 45 min).

^{1}H-NMR (400 MHz, D_2O): δ/ppm = 8.53 (s, 1H, C6-H), 8.51 (s, 1H, C6*-H), 5.96 (d, $^3J_{H–F}$ = 17.7 Hz, 1H, C1'-H), 5.96 (d, $^3J_{H–F}$ = 17.3 Hz, 1H, C1'*-H), 5.18-5.00 (m, 2H, C2'-H, C2'*-H), 4.35-4.21 (m, 2H, C3'-H, C3'*-H), 4.18-4.11 (m, 2H, C4'-H, C4'*-H), 4.03 (dd, ^2J = 13.1 Hz, ^3J = 2.3 Hz, 2H, C5'-H_a, C5'*-H_a)*, 3.82-3.79 (m, 2H, C5'-H_b, C5'*-H_b)* 2.96 (s, 6H, C_m-H_3, C_m*-H_3).

^{13}C-NMR (101 MHz, D_2O): δ/ppm = 161.6 (C4), 161.5 (C4*), 155.3 (C2), 155.3 (C2*), 143.4 (C6), 143.3 (C6*), 107.0 (C5), 106.9 (C5*), 93.6 (d, $^1J_{C–F}$ = 185.1 Hz, C2'), 93.5 (d, $^1J_{C–F}$ = 185.2 Hz, C2'*), 89.7 (d, $^2J_{C–F}$ = 34.9 Hz, C1'), 89.6 (d, $^2J_{C–F}$ = 34.7 Hz, C1'*), 82.3 (C4'),

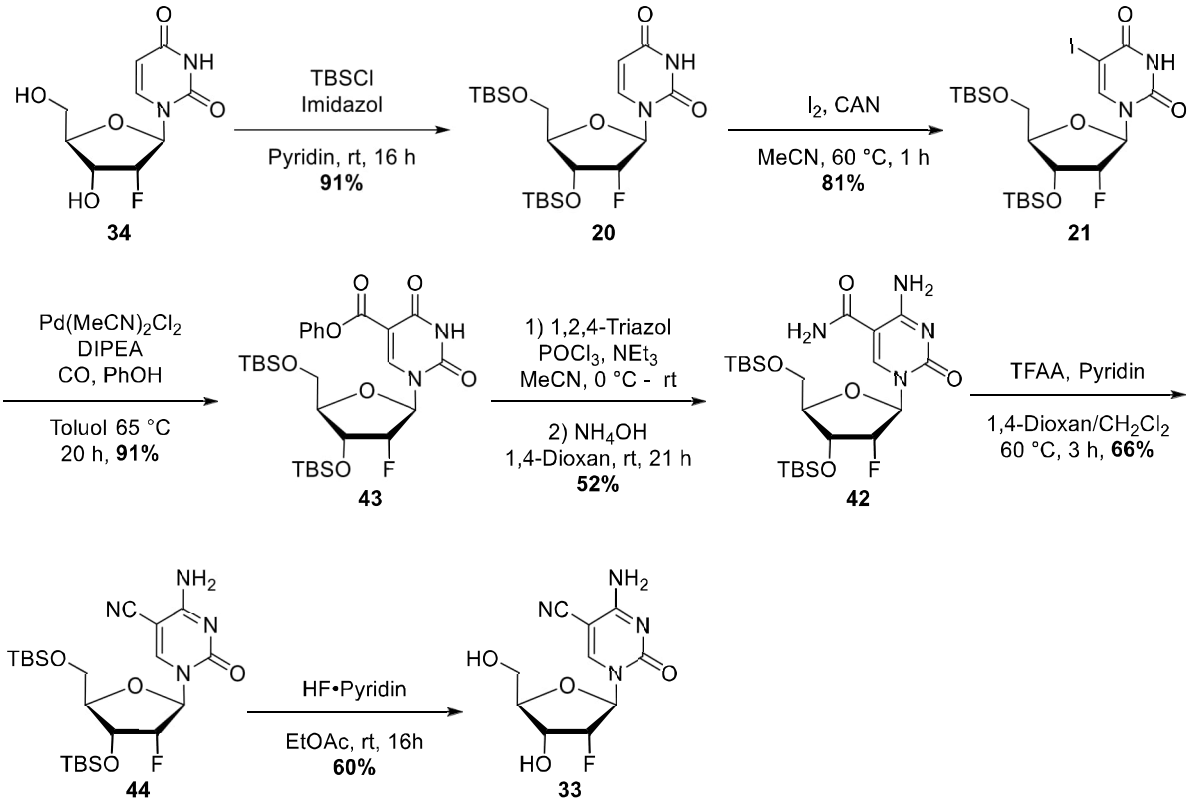
82.2 (C4'*), 67.1 (d, $^{2}J_{C-F}$ = 16.6 Hz, C3'), 67.0 (d, $^{2}J_{C-F}$ = 16.6 Hz, C3'*), 58.9 (C5'), 58.8 (C5'*), 38.6 (C_m), 38.6 (C_m*).

^{19}F-NMR (377 MHz, D_2O): δ/ppm = −196.4, −207.6 (m, C2'-*F*).

HRMS (ESI): berechnet für $C_{10}H_{15}FN_3O_5S$ $[M+H]^+$ 308.0711, gefunden: 308.0713.

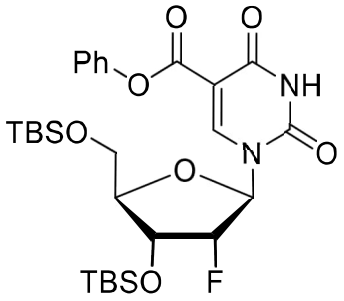
IR (ATR): v/cm^{-1} = 3318 (bw), 3065 (bw), 2954 (w); 1693 (s), 1623 (w), 1471 (w), 1453 (m), 1408 (w), 1361 (w), 1254 (m), 1074 (s), 995 (w), 979 (w), 874 (m), 776 (m), 703 (w), 670 (w).

5.2.7 Syntese von 5-(Cyano)-2'-desoxy-2'-(*R*)-fluoro-cytidin (33)



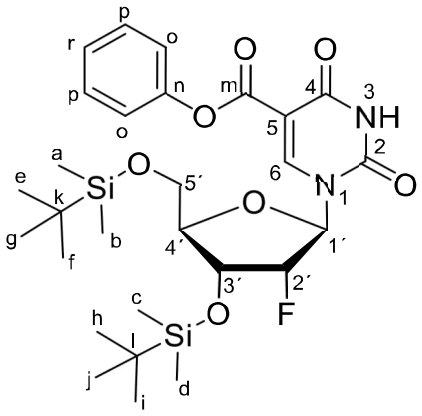
Schema 5.6: Syntheseschema des 2'-F-fdC Analogon 5-CN-2'-F-dC (**33**) ausgehend von kommerziell erwärblichem 2'F-dU (**34**).

5.2.7.1 3',5'-Bis-*O*-(*tert*-butyldimethylsilyl)-5-(phenoxycarbonyl)-2'-desoxy-2'-(*R*)-fluoro-uridin (43)



43

Nach der *ALLGEMEINEN REATIONSVORSCHRIFT II* wurde das 5-iodierte Nukelosid **21** (2.50 g, 4.16 mmol, 1.0 Äq.) mit $[Pd(MeCN)_2Cl_2]$ (54 mg, 0.2 mmol, 5.0 Mol%), Phenol (1.96 g, 20.8 mmol, 5.0 Äq.) und DIPEA (1.08 g, 1.45 mL, 8.32 mmol, 2.0 Äq.) in Toluol (60 mL) umgesetzt und 18 h bei 65 °C und 3.5 bar CO gerührt. Aufeinigung mittels Säulenchromatographie (SiO_2, *i*Hexan/EtOAc = 4:1) ergab die carboxylierte Verbindung **43** (2.25 g, 3.78 mmol, 91 %) in Form eines farblosen Feststoffes.



R_f = 0.40 (SiO_2, *i*Hexan/EtOAc= 2:1).

^{1}H-NMR (400 MHz, $CDCl_3$): δ/ppm = 8.57 (s, 1H, C6-H), 8.14 (s, 1H, N3-H), 7.44-7.36 (m, 2H, C_o-H)*, 7.23-7.27 (m, 1H, C_m-H), 7.18-7.12 (m, 2H, C_p-H)*, 6.05 (dd, $^3J_{H–F}$ = 14.8 Hz, ^3J = 4.0 Hz, 1H, C1'-H), 4.99 (dt, $^2J_{H–F}$ = 53.0 Hz, ^3J = 4.4 Hz, 1H, C2'-H), 4.36 (dt, $^3J_{H–F}$ = 11.3 Hz, ^3J = 5.0 Hz, ^3J = 4.3 Hz, 1H, C3'-H), 4.11–4.04 (m, 1H, C4'-H), 3.94 (dd, ^2J = 11.8 Hz, ^3J = 2.3 Hz, 1H, C5'-H_a)**, 3.77 (dd, ^2J = 11.8 Hz, ^3J = 3.1 Hz, 1H, C5'-H_b)**, 0.92 (s, 9H, C_{e-g}-H_3)***, 0.86 (s, 9H, C_{h-j}-H_3)***, 0.14 (s, 3H, C_a-H_3)****, 0.12 (s, 3H, C_b-H_3)****, 0.06(s, 3H, C_c-H_3)****, 0.04 (s, 3H, C_d-H_3)****.

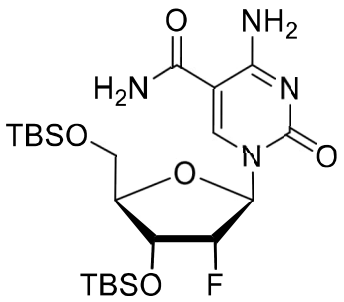
^{13}C-NMR (101 MHz, $CDCl_3$): δ/ppm = 160.8 (C4)*, 157.9 (C_m)*, 150.3 (C_n), 148.7 (C2), 148.4 (C6), 129.4 (2C, C_o)**, 126.1 (C_r), 121.7 (2C, C_p)**, 105.6 (C5), 92.2 (d, $^1J_{C–F}$ = 195.8 Hz, C2'), 89.1 (d, $^2J_{C–F}$ = 33.7 Hz, C1'), 85.7 (C4'), 70.0 (d, $^2J_{C–F}$ = 15.4 Hz, C3'), 61.9 (C5'), 25.9 (3C, C_{e-g})***, 25.6 (3C, C_{h-j})***, 18.4 (C_k)****, 18.1 (C_l)****, −4.7 (d, C_a)*****, −5.1 (C_b)*****, −5.4 (C_c)*****, −5.6 (C_d)*****.

^{19}F-NMR (377 MHz, $CDCl_3$): δ/ppm = −204.4 (dt, $^2J_{F–H}$ = 52.6 Hz, $^3J_{F–H}$ = 11.4 Hz, C2'-*F*).

HRMS (ESI): berechnet für $C_{28}H_{44}FN_2O_7Si_2$ $[M+H]^+$ 595.2671, berechnet: 595.2672.

IR (ATR): v/cm^{-1} = 3171 (bw), 3074 (bw), 2953 (w), 2930 (w), 2892 (w), 2857 (w), 1767 (s), 1713 (s), 1690 (s), 1626 (bw), 1455 (m), 1430 (w), 1388 (w), 1361 (w), 1334 (w), 1296 (w), 1277 (w), 1262 (m), 1190 (s), 1164 (m), 1127 (w), 1098 (m), 1085 (m), 1074 (m), 1040 (m), 1006 (w), 963 (w). 937 (w), 910 (w), 864 (m), 835 (vs), 777 (vs), 756 (m), 688 (m), 673 (w), 658 (m).

5.2.7.2 3',5'-Bis-*O*-(*tert*-butyldimethylsilyl)-5-(carboxamido)-2'-desoxy-2'-(*R*)-fluoro-cytidin (42)

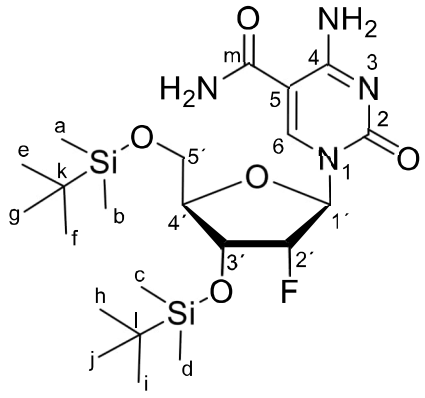


42

1,2,4-Triazol (0.73 g, 10.6 mmol, 9.0 Äq.) wurde in MeCN (30 mL) gelöst, auf 0 °C gekühlt und 10 min bei 0 °C gerührt. $POCl_3$ (0.22 mL, 2.35 mmol, 2.0 Äq.) wurde langsam zugetropft und die Reaktionsmischung wurde weitere 10 min bei 0 °C gerührt. Anschließend wurde NEt_3 (1.4 mL, 10.1 mmol, 8.6 Äq.) tropfenweise hinzugegeben und die Mischung wurde 20 min bei 0 °C gerührt. Im nächsten Schritt wurde eine Lösung von Verbindung **43** (0.70 g, 1.2 mmol, 1.0 Äq.) in MeCN (5.0 mL) langsam zugetropft. Die Reaktionsmischung wurde 20 h bei Raumtemperatur gerührt und anschließend durch Zugabe einer gesättigten wässrigen NH_4Cl-Lösung (50 mL) neutralisiert. MeCN wurde unter vermindertem Druck entfernt und die Lösung mit EtOAc (3 x 100 mL) extrahiert. Die vereinigten organischen Phasen wurden mit einer

gesättigten, wässrigen NaCl-Lösung (2 x 100 mL) gewaschen und über Na_2SO_4 getrocknet und die Lösungsmittel wurden *in vacuo* entfernt.

Das erhaltene Triazolderivat wurde in 1,4-Dioxan (2.5 mL) gelöst und tropfenweise mit einer wässrigen Lösung von NH_3 (25%, 6.3 mL) versetzt. Nachdem die Reaktionsmischung 21 h bei Raumtemperatur gerührt wurde, wurde sie durch Zugabe einer gesättigten wässrigen NH_4Cl-Lösung (35 mL) beendet und mit EtOAc (3 x 50 mL) extrahiert. Die vereinigten organischen Phasen wurden mit einer gesättigten wässrigen NaCl-Lösung (200 mL) gewaschen und anschließend über Na_2SO_4 getrocknet. Lösungsmittel wurden *in vacuo* entfernt und das erhaltene Rohprodukt wurde säulenchromatographisch aufgereinigt (SiO_2, CH_2Cl_2/MeOH = 20:1) um die Verbindung **42** in Form eines farblosen Feststoffes (0.32 g, 0.62 mmol, 52%) zu erhalten.



R_f = 0.50 (SiO_2, CH_2Cl_2/MeOH= 10:1).

^{1}H-NMR (400 MHz, $CDCl_3$): δ/ppm = 8.32 (s, 1H, C6-H), 5.93 (dd, $^3J_{H–F}$ = 17.1 Hz, ^3J = 1.4 Hz, 1H, C1'-H), 5.54 (s, 3H, NH_2), 5.02 (dd, $^2J_{H–F}$ = 52.1 Hz, ^3J = 4.4 Hz, 1H, C2'-H), 4.15–4.25 (m, 2H, C3'-H, C4'-H), 4.11 (dd, ^2J = 11.7 Hz, ^3J = 1.7 Hz, 1H, C5'-H_a)*, 3.83 (dd, ^2J = 11.7 Hz, ^3J = 2.9 Hz, 1H, C5'-H_b)*, 0.92 (s, 9H, $C_{e\text{-}g}$-H_3)**, 0.90 (s, 9H, $C_{h\text{-}j}$-H_3)**, 0.13 (s, 3H, C_a-H_3)***, 0.12 (s, 3H, C_b-H_3)***, 0.10 (s, 3H, C_c-H_3)***, 0.09 (s, 3H, C_d-H_3)***.

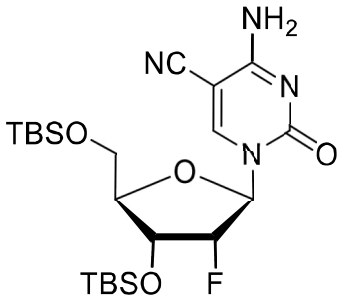
^{13}C-NMR (101 MHz, $CDCl_3$): δ/ppm = 166.9 (C2)*, 163.7 (C_m)*, 153.7 (C4), 143.5 (C6), 98.2 (C5), 92.5 (d, $^1J_{C–F}$ = 191.8 Hz, C2'), 90.1 (d, $^2J_{C–F}$ = 34.4 Hz, C1'), 84.0 (C4'), 68.5 (d, $^2J_{C–F}$ = 15.4 Hz, C3'), 61.3 (C5'), 26.0 (3C, $C_{e\text{-}g}$)**, 25.5 (3C, $C_{h\text{-}j}$)**, 18.6 (C_k)***, 18.0 (C_l)***, −4.5 (d, C_a)****, −5.0 (C_b)****, −5.1(C_c)****, −5.4 (C_d)****.

^{19}F-NMR (377 MHz, $CDCl_3$): δ/ppm = −200.0 (dt, $^2J_{F–H}$ = 51.4 Hz, $^3J_{F–H}$ = 18.5 Hz, C2'-*F*).

HRMS (ESI): berechnet für $C_{22}H_{42}FN_4O_5Si_2$ $[M+H]^+$ 517.2678, berechnet: 517.2675.

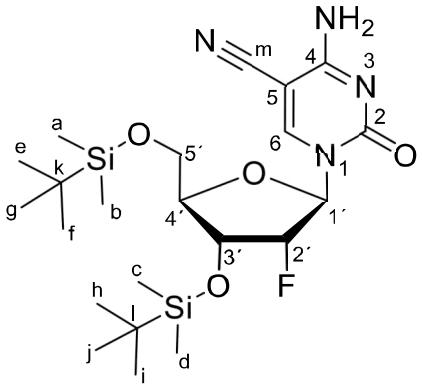
IR (ATR): v/cm^{-1} = 3315 (w), 3192 (bw), 3069 (bw), 2954 (w), 2929 (w), 2903 (w), 2857 (w), 1697 (s), 1622 (m), 1471 (m), 1453 (m), 1408 (w), 1391 (w), 1361 (w), 1323 (w), 1302 (w), 1253 (m), 1224 (w), 1156 (w), 1120 (m), 1074 (m), 995 (w), 979 (w), 938 (w), 874 (w), 857 (m), 833 (vs), 779 vs), 702 (w), 671 (m).

5.2.7.3 3',5'-Bis-*O*-(*tert*-butyldimethylsilyl)-5-(cyano)-2'-desoxy-2'-(*R*)-fluoro-cytidin (44)



44

Nukleosid **42** (0.20 g, 0.40 mmol, 1.0 Äq.) wurde in einem Gemisch von 1,4-Dioxan/CH_2Cl_2 (3 mL/2 mL) gelöst, auf 0 °C abgekühlt und mit Pyridin (0.12 g, 0.13 mL, 1.55 mmol, 4.0 Äq.) und Trifluoroessigsäureanhydrid (0.18 g, 0.12 mL, 0.85 mmol, 2.20 Äq.) in 1,4-Dioxan (1 mL) versetzt. Die Reaktionslösung wurde 3 h bei Raumtemperatur gerührt und anschließend mit einer gesättigten, wässrigen $NaHCO_3$-Lösung (100 mL) verdünnt. Die wässrige Phase wurde mit CH_2Cl_2 (3 x 80 mL) extrahiert, die vereinigten organsichen Phasen mit einer gesättigten, wässrigen NaCl-Lösung (200 mL) gewaschen und über Na_2SO_4 getrocknet. Nachdem Entfernen der flüchtigen Bestandteile unter vermindertem Druck wurde das Rohprodukt säulenchromatographisch (SiO_2, CH_2Cl_2/MeOH = 30:1) aufgereinigt. Die cyano-substituirte Verbindung **44** (0.13 g, 0.26 mmol, 66%) wurde in Form eines Farblosen Feststoffes erhalten.



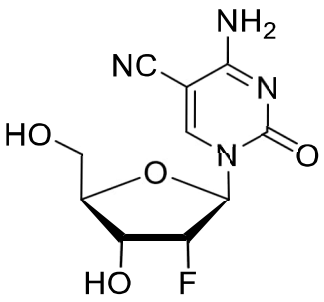
R_f = 0.60 (SiO_2, CH_2Cl_2/MeOH= 20:1).

^{1}H-NMR (400 MHz, $CDCl_3$): δ/ppm = 8.37 (s, 1H, C6-H), 6.02 (d, $^3J_{H–F}$ = 14.5 Hz, 1H, C1'-H), 4.92 (dd, $^2J_{H–F}$ = 51.6 Hz, ^3J = 3.9 Hz, 1H, C2'-H), 4.09-4.26 (m, 2H, C3'-H, C4'-H), 4.16 (dd, ^2J = 11.9 Hz, ^3J = 1.8 Hz, 1H, C5'-H_a)*, 3.83 (dd, ^2J = 11.9 Hz, ^3J = 1.2 Hz, 1H, C5'-H_b)*, 0.98 (s, 9H, C_{e-g}-H_3)**, 0.90 (s, 9H, C_{h-j}-H_3)**, 0.19 (s, 3H, C_a-H_3)***, 0.18 (s, 3H, C_b-H_3)***, 0.13 (s, 3H, C_c-H_3)***, 0.11 (s, 3H, C_d-H_3)***.

HRMS (ESI): berechnet für $C_{22}H_{40}FN_4O_4Si_2$ $[M+H]^+$ 499.2572, berechnet: 499.2572.

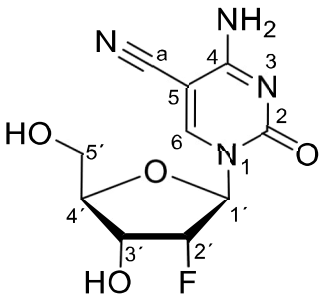
IR (ATR): v/cm^{-1} = 3069 (bw), 2954 (w), 2929 (w), 2892 (w), 2857 (w), 2148 (w), 1700 (vs), 1653 (w), 1622 (w), 1559 (w), 1540 (w), 1507 (w), 1472 (w), 1456 (w), 1361 (w); 1255 (m), 1157 (w), 1121 (m), 1074 (m), 980 (w), 836 (vs), 781 (s), 723 (w), 672 (w), 657 (w).

5.2.7.4 5-(Cyano)-2'-desoxy-2'-(*R*)-fluoro-cytidin (33)



33

Nach der *ALLGEMEINEN REAKTIONSVORSCHRIFT I* wurde **44** (0.12 g, 0.24 mmol, 1.00 Äq.) in EtOAc (3.0 mL) mit HF • Pyridin (70%, 94.0 µL, 3.62 mmol, 15.0 Äq.) versetzt. Die Reaktionsmischung wurde 20 h bei Raumtemperatur gerührt, anschließend auf 0 °C gekühlt, mit TMS • OMe (0.81 g, 1.10 mL, 7.71 mmol, 32.0 Äq.) versetzt und für 1 h bei 0 °C gerührt. Aufreinigung mittels semipräparativer HPL-Chromatographie (MeCN/H_2O, MeCN: 0 → 20%, 45 min, Retentionszeit: 30 min) ergab Verbindung **33** (38.0 mg, 0.14 mmol, 60%) in Form eines farblosen Pulvers.



Retentionszeit = 30 min (C_{18}-SiO_2, MeCN/H_2O, MeCN: 0 → 20%, 45 min).

^{1}H-NMR (400 MHz, D_2O): δ/ppm = 8.72 (s, 1H, C6-H), 6.00 (d, $^3J_{H–F}$ = 17.6 Hz, 1H, C1'-H), 5.14 (dd, $^2J_{H–F}$ = 52.3 Hz, ^3J = 4.3 Hz, 1H, C2'-H), 4.31 (ddd, $^3J_{H–F}$ = 23.7 Hz, ^3J = 4.3 Hz, ^3J = 9.3 Hz 1H, C3'-H), 4.21 (d, ^3J = 9.3 Hz 1H, C4'-H), 4.09 (dd, ^2J = 13.2 Hz, ^3J = 2.3 Hz, 1H, C5'-H_a)*, 3.88 (dd, ^2J = 13.2 Hz, ^3J = 3.7 Hz, 1H, C5'-H_b)*.

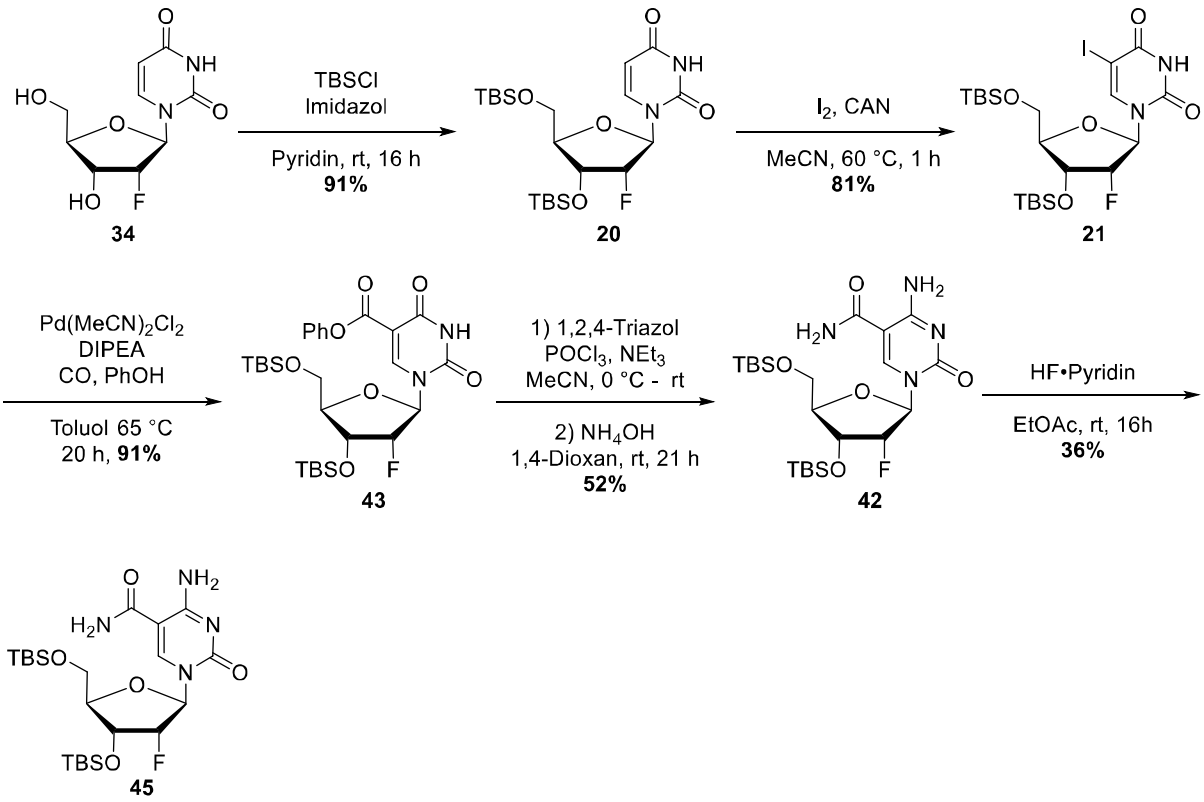
^{13}C-NMR (101 MHz, D_2O): δ/ppm = 163.4 (C2), 154.9 (C4)*, 150.8 (C6), 114.1 (C5)*, 93.6 (d, $^1J_{C–F}$ = 185.5 Hz, C2'), 89.9 (d, $^2J_{C–F}$ = 35.5 Hz, C1'), 82.5 (C4'), 81.3 (C_a), 67.3 (d, $^2J_{C–F}$ = 16.9 Hz, C3'), 59.2 (C5').

^{19}F-NMR (377 MHz, D_2O): δ/ppm = −201.7 (dt, $^2J_{F–H}$ = 52.3 Hz, $^3J_{F–H}$ = 23.6 Hz, C2'-*F*).

HRMS (ESI): berechnet für $C_{10}H_{12}FN_4O_4$ $[M+H]^+$ 271.0843, berechnet: 271.0838.

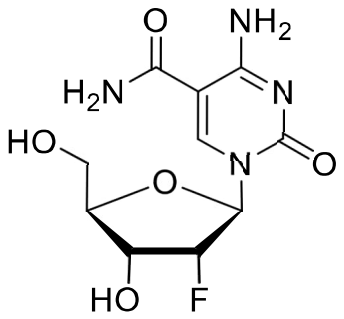
IR (ATR): ν/cm^{-1} = 3335 (bw), 3211 (bw), 2927 (w), 2891 (w), 2227 (w), 1646 (vs), 1493 (m), 1408 (w), 1361 (w); 1295 (w), 1251 (w), 1097 (s), 1059 (m), 984 (w), 949 (w), 916 (w), 853 (w), 782 (m), 760 (w), 723 (w), 705 (w), 672 (w).

5.2.8 Syntese von 5-(Carboxamido)-2'-desoxy-2'-(*R*)-fluoro-cytidin (45)



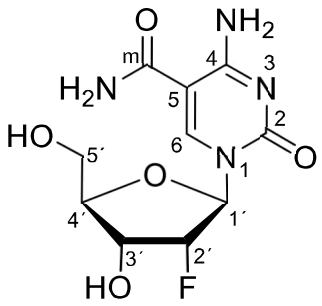
Schema 5.7: Syntheseschema des 2'-F-cadC Analogon 5-$CONH_2$-2'-F-dC (**45**) ausgehend von kommerziell erwärblichem 2'F-dU (**34**).

5.2.8.1 5-(Carboxamido)-2'-desoxy-2'-(*R*)-fluoro-cytidin (45)



45

Nach der *ALLGEMEINEN REAKTIONSVORSCHRIFT I* wurde **42** (43.0 mg, 83.0 µmol, 1.0 Äq.) in EtOAc (1.5 mL) mit HF • Pyridin (70%, 22.0 µL, 12.4 µmol, 15.0 Äq.) versetzt. Die Reaktionsmischung wurde 18 h bei Raumtemperatur gerührt, anschließend auf 0 °C gekühlt, mit TMS • OMe (0.27 g, 0.37 mL, 0.26 mmol, 32.0 Äq.) versetzt und für 1 h bei 0 °C gerührt. Aufreinigung mittels semipräparativer HPL-Chromatographie (MeCN/H_2O, MeCN: 0 → 20%, 45 min, Retentionszeit: 18 min) ergab Verbindung **45** (8.5 mg, 29.5 µmol, 36%) in Form eines farblosen Pulvers.



Retentionszeit = 18 min (C_{18}-SiO_2, MeCN/H_2O, MeCN: 0 → 20%, 45 min).

^{1}H-NMR (400 MHz, D_2O): δ/ppm = 8.30 (s, 1H, C6-H), 5.88 (dd, $^3J_{H–F}$ = 17.0 Hz, ^3J = 1.2 Hz, 1H, C1'-H), 5.10 (dd, $^2J_{H–F}$ = 51.8 Hz, ^3J = 4.3 Hz, 1H, C2'-H), 4.15–4.25 (m, 2H, C3'-H, C4'-H), 4.11 (dd, ^2J = 11.8 Hz, ^3J = 1.5 Hz, 1H, C5'-H_a)*, 3.83 (dd, ^2J = 11.8 Hz, ^3J = 2.4 Hz, 1H, C5'-H_b).

^{13}C-NMR (101 MHz, D_2O): δ/ppm = 166.9 (C2)*, 163.7 (C_m)*, 153.7 (C4), 143.5 (C6), 98.2 (C5), 92.5 (d, $^1J_{C–F}$ = 191.8 Hz, C2'), 90.1 (d, $^2J_{C–F}$ = 34.4 Hz, C1'), 84.0 (C4'), 68.5 (d, $^2J_{C–F}$ = 15.4 Hz, C3'), 61.3 (C5').

^{19}F-NMR (377 MHz, D_2O): δ/ppm = −200.8 (dt, $^{2}J_{F-H}$ = 52.0 Hz, $^{3}J_{F-H}$ = 24.1 Hz, C2'-*F*).

HRMS (ESI): berechnet für $C_{10}H_{14}FN_4O_5$ $[M+H]^+$ 289.0943, gefunden: 289.0947.

IR (ATR): v/cm^{-1} = 3334 (bw), 3196 (bw), 2953 (w), 2930 (w), 2894 (w), 2858 (w), 1698 (m), 1656 (s), 1639 (s), 1591 (w), 1492 (m), 1398 (m), 1306 (m), 1125 (s), 1077 (s9; 1006 (w), 833 (vs), 775 (vs), 671 (m).

5.3 Material und Methoden biochemische Experimente

5.3.1 Zellbiologische Verfütterungsexperimente, DNA und Nukleosid Pool Isolation

Zelllinien

Eine detaillierte Beschreibung der verwendeten Zelllinien für die in dieser Arbeit präsentierten Experimente ist in der Dissertation von *Angie Kirchner* und *Ewelina Kaminska* zu finden.

Zellernte und Isolierung genomischer DNA

Bevor die Zellen geerntet wurden, wurde sie mit einem PBS-Puffer (*Merck*) gewaschen und durch Zugabe von RLT buffer (*Qiagen*) mit 400µM 2,6-di-tert-butyl-4-methylphenol (BHT) und Desferoxamin-mesylat (DM) erfolgte die Lyse der entsprechenden Zellen.
Im nächsten Schritt wurden die Proben wie folgt behandelt: In 2 mL Mikrozentrifugen Gefäßen (*Eppendorf*) wurde die Lösung mit edelstahl Kugel mit einem 5 mm Durchmesser (eine pro Probe) in einer MM400 Mühle (*Retsch*) bei 30 Hz für 1 Minute behandelt und anschließend für 5 Minuten bei 21000 rcf zentrifugert. Genomische DNA (gDNA) wurde mit *Zymo Quick gDNA mini-prep® kit* (*Zymo research*) extrahiert. Die Probe wurde mit RNAse A (*Qiagen*) bei 0.2 mg/mL und *Genomic Lysis Buffer* versetzt. Ein Teil der Probe (400 µL) wurde 15 Minuten auf der Aufreinigungssäule inkubiert. Die gDNA wurde mit DNAse-freiem ddH_2O mit 20µM BHT eluiert. Anschließend wurde die Konzentration der erhaltenen Proben am Nanodrop N60 (*Implen*) bestimmt.

Isolierung und Aufreinigung vom löslichen Nukleosid Pool

Die Zellen wurden mit Hilfe von TrypLE Reagenz (*Gibco*) trypsinisiert und anschließend mit Tryptan Blau am Zellzähler *Countess II FL Automated Cell Counter* (*Life Technologies, ThermoFisher Scienitfic*) gezählt. Die Probe wurde bei 260 G für 3 Minuten zentrifugiert. Das Pellet wurde in einer eiskalten Mischung (H_2O/MeCN 1:1, 1 mL pro $1x10^6$ Zellen) aufgenommen, leicht gevortext, für 15 min inkubiert und 10 min bei 4 °C und 16000 G zentrifugiert. Der Nukleosid Pool erhaltendte Überstand wurde lyophilisiert und das Pellet mit dem Zelllyse Puffer zur gDNA Isolierung behandelt.

Proben wurde in H_2O (pH = 4, 100 µL H_2O für Nukleosid Pool aus 400 000 Zellen) gelöst. *Supel-Select HBL 60 mg* Säulen (*Merck*) wurde mit 1.5 mL MeOH und 3 mL H_2O (pH = 4) equilibriert. Proben wurde auf die Säule aufgetragen und mit 4 mL H_2O (pH = 4) gewaschen. Säulen wurden unter Hochvakuum für 15 Minuten getrocknet. Anschließend wurden die

Nukleoside mit 1.5 mL MeOH/MeCN = 1:1 eluiert. Eluent als auch die Waschlösung wurden lyophilisiert, in H_2O aufgenommen, die ensprechenden isotopenmarkierten Standards wurden hinzugegeben, die Probe wurde filtriert (über 0.2 µm Filter) und mit UHPLC-MS/MS analog zur Analyse von gDNA analysiert.

5.3.2 Enzymatischer DNA Verdau

1 – 10 µg genomischer DNA in 35 µL H_2O wurden nach folgendem Protokoll vedaut: 2 µL Degradase plus (*Zymo research*), 5 µL of Degradase plus Reaktionspuffer (*Zymo research*), 0.5 µL of Benzonase nuclease (*Merck*, *Serratia marcescens*) und eine spezifische Menge von isotopen-markierten internen Standards wurden hinzugefügt. Die Probe wurde für 4 h bei 37 °C inkubiert. Falls nicht direkt für die MS-Analyse verwendet, wurden die Proben bei -20 °C gelagert. Vor der LC-MS/MS Analyse, wurden die Proben mit einer *AcroPrep Advance* Filterplatte 0.2 µm *Supor* (*Pall Life Science*) filtriert.

5.3.3 LC-MS/MS Analyse von DNA-Proben

Für die quantitative UHPLC-MS/MS Analyse von verdauten DNA Proben wurde ein *Agilent 1290 UHPLC System* mit einem UV detector und *Agilent 6490 triple quadrupole Massenspektrometer* verwendet. Die zu analysierenden Nukleoside wurde mit der *stable isotope dilution technique*[203] (Stabilen-isotopenverdünnungstechnik) exakt quantifiziert. Die quellenabhängigen Parameter waren die folgenden: Gas Temperatur 80 °C, Gas Fluss 15 L/min (N_2), Nebulizer 30 psi, sheath Gas Heater 275 °C, sheath Gas Fluss 11 L/min (N_2), Kapillarenspannung 2.500 V im positiven Ionisierungsmodus, Kapillarenspannung −2.250 V im negativen Ionisierungsmodusd und nozzle voltage 500 V. Die Fragmentor Spannung betrug 380 V/ 250 V. Delta EMV betrug 500 (positiv Modus) and 800 (negativ Modus). Die Verbindungsabhängigen Parameter sind in **Tabelle 5.1** bis **Tabelle 5.6** aufgelistet. Stationäre Phase für die UHPL-Chromatographie war eine *Poroshell 120 SB-C8 Säule* (*Agilent,* 2.7 µm, 2.1 mm × 150 mm) bei 35 °C Säulenofentemperatur. Als mobile Phase wurde ein Gradient von Wasser und Acetonitril, mit 0.0085% (v/v) Ameisensäure, bei einer Flussrate von 0.35 mL/min: 0 →4 min; 0 →3.5% (v/v) MeCN; 4 →7.9 min; 3.5 →5% MeCN; 7.9 →8.2 min; 5 →80% MeCN; 8.2 →11.5 min; 80% MeCN; 11.5 →12 min; 80 →0% MeCN; 12 →14 min; 0% MeCN verwendet. Zwischen 0 – 1.5 min und nach 12 min wurde das Effluent mittels eines Valco Ventils in Lösemittelabfall geleitet. Der Autosampler war auf 4 °C gekühlt. Das Injektionsvolumen der verdauten DNA Proben betrug 39 µL.

Tabelle 5.1: Substanzabhängigen LC-MS/MS Parameter zur Quantifizierung der markierten Nukleoside in genomischer DNA aus den **Verfütterungsexperimenten mit $^{13}C_5$-$^{15}N_2$-fdC** (**1**). *CE: collision energy, Kollisionsenergie, CAV: collision cell accelerator voltage,* Die Nukleoside wurden als positiv- ($[M+H]^+$ Intermediat) als auch im negativ-Modus ($[M-H]^-$ Intermediat) detektiert im *multiple reaction monitoring Modus (MRM).*

Verbindung	*Precursor ion (m/z)*	*MS1 Auflösung*	*Produkt Ion (m/z)*	*MS2 Auflösung*	*Dwell time [ms]*	*CE (V)*	*CAV (V)*	*Polarität*
Time segment 1.5-3.3 min								
[$^{13}C_5$-$^{15}N_2$]-cadC	279.10	Wide	158.03	Wide	50	5	5	Positiv
[$^{15}N_2$]-cadC	274.08	Wide	158.03	Wide	50	5	5	Positiv
cadC	272.09	Wide	156.04	Wide	20	5	5	Positiv
[$^{13}C_5$-$^{15}N_2$]-mdC	249.12	Wide	128.06	Wide	60	60	1	Positiv
[D_3]-mdC	245.13	Wide	129.09	Wide	60	60	1	Positiv
mdC	242.11	Wide	126.07	Wide	50	60	1	Positiv
[$^{13}C_5$-$^{15}N_2$]-dC	235.11	Wide	114.04	Wide	60	10	2	Positiv
[$^{13}C_9$-$^{15}N_3$]-dC	240.1	Wide	119.1	Wide	80	5	2	Positiv
[$^{13}C_5$-$^{15}N_2$]-dU	234.08	Wide	185.06	Wide	15	5	5	Negativ
[$^{15}N_2$]-dU	229.06	Wide	185.06	Wide	15	5	5	Negativ
Time segment 4-9 min								
[$^{13}C_5$-$^{15}N_2$]-fdC	263.1	Wide	142.04	Wide	75	5	5	Positiv
[$^{15}N_2$]-fdC	258.09	Wide	142.04	Wide	75	5	5	Positiv
fdC	256.09	Wide	140.05	Wide	10	5	5	Positiv
[$^{13}C_5$-$^{15}N_2$]-fdU	262.07	Wide	141.01	Wide	80	15	5	Negativ
[$^{15}N_2$]-fdU	257.06	Wide	141.01	Wide	80	15	5	Negativ
fdU	255.06		139.01	Wide	10	15	5	Negativ
[$^{13}C_5$-$^{15}N_2$] -dT	250.11		129.04	Wide	10	8	5	Positiv
[^{13}C-$^{15}N_2$]-dT	246.12		130.07	Wide	10	8		Positiv

dT	243.10	Wide	127.05	Wide	10	8	5	Positiv
[$^{13}C_5$-$^{15}N_2$]-dU	234.08	Wide	185.06	Wide	15	5	5	Negativ
[$^{15}N_2$]-dU	229.06	Wide	185.06	Wide	15	5	5	Negativ

Tabelle 5.2: Substanzabhängigen LC-MS/MS Parameter zur Quantifizierung der fluorierten Nukleoside in genomischer DNA aus den **Verfütterungsexperimenten mit 2'-F-cadC *prodrugs* (17, 18, 19**). *CE: collision energy, Kollisionsenergie, CAV: collision cell accelerator voltage, Die Nukleoside wurden als positiv- ([M+H]$^+$ Intermediat) als auch im negativ-Modus ([M-H]$^-$ Intermediat) detektiert im multiple reaction monitoring Modus (MRM).*

Verbindung	*Precursor ion (m/z)*	*MS1 Auflösung*	*Produkt Ion (m/z)*	*MS2 Auflösung*	*Dwell time [ms]*	*CE (V)*	*CAV (V)*	*Polarität*
Time segment 1.5-3.7 min								
F-dC	246.09	Wide	112.06	Wide	70	15	3	Positiv
[$^{15}N_2$]-F-dC	248.08	Wide	114.04	Wide	70	15	3	Positiv
[$^{15}N_2$]-cadC	274.08	Wide	158.03	Wide	40	5	5	Positiv
cadC	272.09	Wide	156.04	Wide	40	5	5	Positiv
F-hmdC	276.10	Wide	142.06	Wide	50	10	3	Positiv
[$^{15}N_2$]-F-hmdC	278.09	Wide	144.06	Wide	50	10	3	Positiv
[$^{15}N_2$,D_2]-hmdC	262.12	Wide	146.07	Wide	25	27	1	Positiv
hmdC	258.11	Wide	142.06	Wide	25	27	1	Positiv
[D_3]-mdC	245.13	Wide	129.09	Wide	50	60	1	Positiv
mdC	242.11	Wide	126.07	Wide	50	60	1	Positiv
F-cadC	290.08	Wide	156.04	Wide	80	5	5	Positiv
[$^{15}N_2$]-F-cadC	292.08	Wide	158.04	Wide	80	5	5	Positiv
Time segment 3.7-4.7 min								
[D_3]-F-mdC	263.12	Wide	129.09	Wide	80	15	3	Positiv
F-mdC	260.10	Wide	126.07	Wide	80	15	3	Positiv
F-cadC	290.08	Wide	156.04	Wide	80	5	5	Positiv
[$^{15}N_2$]-F-cadC	292.08	Wide	158.04	Wide	80	5	5	Positiv
Time segment 4.7-10 min								

F-CO_2Me-dC	304.09	Wide	170.06	Wide	70	10	3	Positiv
F-CO_2Et-dC	318.11	Wide	184.07	Wide	70	10	3	Positiv
F-CO_2Bn-dC	380.13	Wide	246.09	Wide	70	10	3	Positiv
F-fdC	274.08	Wide	140.05	Wide	30	15	3	Positiv
[$^{15}N_2$]-F-fdC	276.08	Wide	142.04	Wide	30	15	3	Positiv
F-fdU	273.05	Wide	253.05	Wide	30	3	5	Negativ
[$^{15}N_2$]-fdC	258.09	Wide	142.04	Wide	30	5	5	Positiv
fdC	256.09	Wide	140.05	Wide	30	5	5	Positiv
[$^{15}N_2$]-fdC	142.04	Wide	98.04	Wide	20	13	7	Positiv

Tabelle 5.3: Substanzabhängigen LC-MS/MS Parameter zur Quantifizierung der fluorierten Nukleoside in genomischer DNA aus den **Verfütterungsexperimenten mit 2'-F-COMe-dC** (**31**). *CE: collision energy, Kollisionsenergie, CAV: collision cell accelerator voltage, Die Nukleoside wurden als positiv- ($[M+H]^+$ Intermediat) als auch im negativ-Modus ($[M-H]^-$ Intermediat) detektiert im multiple reaction monitoring Modus (MRM).*

Verbindung	*Precursor ion (m/z)*	*MS1 Auflösung*	*Produkt Ion (m/z)*	*MS2 Auflösung*	*Dwell time [ms]*	*CE (V)*	*CAV (V)*	*Polarität*
Time segment 1.5-3.7 min								
F-dC	246.09	Wide	112.06	Wide	70	15	3	Positiv
[$^{15}N_2$]-F-dC	248.08	Wide	114.04	Wide	70	15	3	Positiv
[$^{15}N_2$]-cadC	274.08	Wide	158.03	Wide	40	5	5	Positiv
cadC	272.09	Wide	156.04	Wide	40	5	5	Positiv
[$^{15}N_2$,D_2]-hmdC	262.12	Wide	146.07	Wide	25	27	1	Positiv
hmdC	258.11	Wide	142.06	Wide	25	27	1	Positiv
[D_3]-mdC	245.13	Wide	129.09	Wide	50	60	1	Positiv
mdC	242.11	Wide	126.07	Wide	50	60	1	Positiv
F-cadC	290.08	Wide	156.04	Wide	80	5	5	Positiv
[$^{15}N_2$]-F-cadC	292.08	Wide	158.04	Wide	80	5	5	Positiv
Time segment 3.7-4.7 min								
F-5-COMe-dC	288.1	Wide	154.06	Wide	110	10	3	Positiv

[D_3]-F-mdC	263.12	Wide	129.09	Wide	50	15	3	Positiv
F-mdC	260.10	Wide	126.07	Wide	50	15	3	Positiv
F-cadC	290.08	Wide	156.04	Wide	70	5	5	Positiv
[$^{15}N_2$]-F-cadC	292.08	Wide	158.04	Wide	70	5	5	Positiv
F-hmU-dN	275.07		155.06		20	3	3	Negativ
F-dU	245.06		225.06		20	3	3	Negativ
Time segment 4.7-12 min								
F-5-COMe-dC	288.1	Wide	154.06	Wide	110	10	3	Positiv
F-fdC	274.08	Wide	140.05	Wide	90	15	3	Positiv
[$^{15}N_2$]-F-fdC	276.08	Wide	142.04	Wide	90	15	3	Positiv
F-fdU	273.05	Wide	253.05	Wide	10	3	5	Negativ
F-dT	259.07	Wide	239.07	Wide	10	3	5	Negativ
[$^{15}N_2$]-fdC	258.09	Wide	142.04	Wide	20	5	5	Positiv
fdC	256.09	Wide	140.05	Wide	20	5	5	Positiv
[$^{15}N_2$]-fdC	142.04	Wide	98.04	Wide	20	13	7	Positiv
fdC	140.05	Wide	97.04	Wide	20	13	7	Positiv

Tabelle 5.4: Substanzabhängigen LC-MS/MS Parameter zur Quantifizierung der fluorierten Nukleoside in genomischer DNA aus den **Verfütterungsexperimenten mit 2'-F-SOMe-dC** (**32**). *CE: collision energy, Kollisionsenergie, CAV: collision cell accelerator voltage, Die Nukleoside wurden als positiv- ($[M+H]^+$ Intermediat) als auch im negativ-Modus ($[M-H]^-$ Intermediat) detektiert im multiple reaction monitoring Modus (MRM).*

Verbindung	*Precursor ion (m/z)*	*MS1 Auflösung*	*Produkt Ion (m/z)*	*MS2 Auflösung*	*Dwell time [ms]*	*CE (V)*	*CAV (V)*	*Polarität*
Time segment 1.5-3.7 min								
F-dC	246.09	Wide	112.06	Wide	70	15	3	Positiv
[$^{15}N_2$]-F-dC	248.08	Wide	114.04	Wide	70	15	3	Positiv
[$^{15}N_2$]-cadC	274.08	Wide	158.03	Wide	40	5	5	Positiv
cadC	272.09	Wide	156.04	Wide	40	5	5	Positiv
[$^{15}N_2$,D_2]-hmdC	262.12	Wide	146.07	Wide	25	27	1	Positiv
hmdC	258.11	Wide	142.06	Wide	25	27	1	Positiv
[D_3]-mdC	245.13	Wide	129.09	Wide	50	60	1	Positiv

mdC	242.11	Wide	126.07	Wide	50	60	1	Positiv
F-cadC	290.08	Wide	156.04	Wide	80	5	5	Positiv
[$^{15}N_2$]-F-cadC	292.08	Wide	158.04	Wide	80	5	5	Positiv
Time segment 3.7-4.7 min								
F-5-SOMe-dC	308.07	Wide	174.03	Wide	110	10	3	Positiv
[D_3]-F-mdC	263.12	Wide	129.09	Wide	50	15	3	Positiv
F-mdC	260.10	Wide	126.07	Wide	50	15	3	Positiv
F-cadC	290.08	Wide	156.04	Wide	70	5	5	Positiv
[$^{15}N_2$]-F-cadC	292.08	Wide	158.04	Wide	70	5	5	Positiv
F-hmU-dN	275.07	Wide	155.06	Wide	20	3	3	Negativ
F-dU	245.06	Wide	225.06	Wide	20	3	3	Negativ
Time segment 4.7-12 min								
F-5-SOMe-dC	308.07	Wide	174.03	Wide	110	10	3	Positiv
F-fdC	274.08	Wide	140.05	Wide	90	15	3	Positiv
[$^{15}N_2$]-F-fdC	276.08	Wide	142.04	Wide	90	15	3	Positiv
F-fdU	273.05	Wide	253.05	Wide	10	3	5	Negativ
F-dT	259.07	Wide	239.07	Wide	10	3	5	Negativ
[$^{15}N_2$]-fdC	258.09	Wide	142.04	Wide	20	5	5	Positiv
fdC	256.09	Wide	140.05	Wide	20	5	5	Positiv
[$^{15}N_2$]-fdC	142.04	Wide	98.04	Wide	20	13	7	Positiv
fdC	140.05	Wide	97.04	Wide	20	13	7	Positiv

Tabelle 5.5: Substanzabhängigen LC-MS/MS Parameter zur Quantifizierung der fluorierten Nukleoside in genomischer DNA aus den **Verfütterungsexperimenten mit 2'-F-CN-dC** (**33**). *CE: collision energy, Kollisionsenergie, CAV: collision cell accelerator voltage, Die Nukleoside wurden als positiv- ($[M+H]^+$ Intermediat) als auch im negativ-Modus ($[M-H]^-$ Intermediat) detektiert im multiple reaction monitoring Modus (MRM).*

Verbindung	*Precursor ion (m/z)*	*MS1 Auflösung*	*Produkt Ion (m/z)*	*MS2 Auflösung*	*Dwell time [ms]*	*CE (V)*	*CAV (V)*	*Polarität*
Time segment 1.5-3.7 min								
F-dC	246.09	Wide	112.06	Wide	70	15	3	Positiv

[$^{15}N_2$]-F-dC	248.08	Wide	114.04	Wide	70	15	3	Positiv
[$^{15}N_2$]-cadC	274.08	Wide	158.03	Wide	40	5	5	Positiv
cadC	272.09	Wide	156.04	Wide	40	5	5	Positiv
[$^{15}N_2$,D_2]-hmdC	262.12	Wide	146.07	Wide	25	27	1	Positiv
hmdC	258.11	Wide	142.06	Wide	25	27	1	Positiv
[D_3]-mdC	245.13	Wide	129.09	Wide	50	60	1	Positiv
mdC	242.11	Wide	126.07	Wide	50	60	1	Positiv
F-cadC	290.08	Wide	156.04	Wide	80	5	5	Positiv
[$^{15}N_2$]-F-cadC	292.08	Wide	158.04	Wide	80	5	5	Positiv
Time segment 3.7-4.7 min								
F-5-CN-dC	271.08	Wide	137.05	Wide	110	10	3	Positiv
[D_3]-F-mdC	263.12	Wide	129.09	Wide	50	15	3	Positiv
F-mdC	260.10	Wide	126.07	Wide	50	15	3	Positiv
F-cadC	290.08	Wide	156.04	Wide	70	5	5	Positiv
[$^{15}N_2$]-F-cadC	292.08	Wide	158.04	Wide	70	5	5	Positiv
F-hmU-dN	275.07	Wide	155.06	Wide	20	3	3	Negativ
F-dU	245.06	Wide	225.06	Wide	20	3	3	Negativ
Time segment 4.7-12 min								
F-5-CN-dC	271.08	Wide	137.05	Wide	110	10	3	Positiv
F-fdC	274.08	Wide	140.05	Wide	90	15	3	Positiv
[$^{15}N_2$]-F-fdC	276.08	Wide	142.04	Wide	90	15	3	Positiv
F-fdU	273.05	Wide	253.05	Wide	10	3	5	Negativ
F-dT	259.07	Wide	239.07	Wide	10	3	5	Negativ
[$^{15}N_2$]-fdC	258.09	Wide	142.04	Wide	20	5	5	Positiv
fdC	256.09	Wide	140.05	Wide	20	5	5	Positiv
[$^{15}N_2$]-fdC	142.04	Wide	98.04	Wide	20	13	7	Positiv
fdC	140.05	Wide	97.04	Wide	20	13	7	Positiv

Tabelle 5.6: Substanzabhängigen LC-MS/MS Parameter zur Quantifizierung der fluorierten Nukleoside in genomischer DNA aus den **Verfütterungsexperimenten mit 2'-F-CONH$_2$-dC** (**45**). *CE: collision energy, Kollisionsenergie, CAV: collision cell accelerator voltage, Die Nukleoside wurden als positiv- ([M+H)]$^+$ Intermediat) als auch im negativ-Modus ([M-H]$^-$ Intermediat) detektiert im multiple reaction monitoring Modus (MRM).*

Verbindung	*Precursor ion (m/z)*	*MS1 Auflösung*	*Produkt Ion (m/z)*	*MS2 Auflösung*	*Dwell time [ms]*	*CE (V)*	*CAV (V)*	*Polarität*
Time segment 1.5-3.7 min								
F-dC	246.09	Wide	112.06	Wide	70	15	3	Positiv
[$^{15}N_2$]-F-dC	248.08	Wide	114.04	Wide	70	15	3	Positiv
[$^{15}N_2$]-cadC	274.08	Wide	158.03	Wide	40	5	5	Positiv
cadC	272.09	Wide	156.04	Wide	40	5	5	Positiv
[$^{15}N_2$,D_2]-hmdC	262.12	Wide	146.07	Wide	25	27	1	Positiv
hmdC	258.11	Wide	142.06	Wide	25	27	1	Positiv
[D_3]-mdC	245.13	Wide	129.09	Wide	50	60	1	Positiv
mdC	242.11	Wide	126.07	Wide	50	60	1	Positiv
F-cadC	290.08	Wide	156.04	Wide	80	5	5	Positiv
[$^{15}N_2$]-F-cadC	292.08	Wide	158.04	Wide	80	5	5	Positiv
Time segment 3.7-4.7 min								
F-5-CONH$_2$-dC	298.09	Wide	155.05	Wide	110	10	3	Positiv
[D_3]-F-mdC	263.12	Wide	129.09	Wide	50	15	3	Positiv
F-mdC	260.10	Wide	126.07	Wide	50	15	3	Positiv
F-cadC	290.08	Wide	156.04	Wide	70	5	5	Positiv
[$^{15}N_2$]-F-cadC	292.08	Wide	158.04	Wide	70	5	5	Positiv
F-hmU-dN	275.07	Wide	155.06	Wide	20	3	3	Negativ
F-dU	245.06	Wide	225.06	Wide	20	3	3	Negativ
Time segment 4.7-12 min								
F-5-CONH$_2$-dC	298.09	Wide	155.05	Wide	110	10	3	Positiv
F-fdC	274.08	Wide	140.05	Wide	90	15	3	Positiv
[$^{15}N_2$]-F-fdC	276.08	Wide	142.04	Wide	90	15	3	Positiv
F-fdU	273.05	Wide	253.05	Wide	10	3	5	Negativ

F-dT	259.07	Wide	239.07	Wide	10	3	5	Negativ
[$^{15}N_2$]-fdC	258.09	Wide	142.04	Wide	20	5	5	Positiv
fdC	256.09	Wide	140.05	Wide	20	5	5	Positiv
[$^{15}N_2$]-fdC	142.04	Wide	98.04	Wide	20	13	7	Positiv
fdC	140.05	Wide	97.04	Wide	20	13	7	Positiv

5.3.4 Nukleosid Quantifizierung in DNA-Proben

Quantifizierung der Nukleoside aus den DNA-Proben wurde mit Hilfe von vorher aufgenommenen internen Kalibriertgeraden durchgeführt. Dies folgte der in der Literatur beschriebenen Technik der *stable isotope dilution technique for mass spectrometry*.[203] Die verwendeten Kalibiegeraden und deren entsprechenden Gültigkeitsbereiche für die in dieser Arbeit untersuchten Nukleoside sind in **Tabelle 5.7** bis **Tabelle 5.9** gezeigt.

Tabelle 5.7: Substanzabhängigen UV-VIS Geradengleichungen der Quantifizierungskalibriergeraden und die enpsrechenden Gültigkeitsbereiche der Kalibriergerade für die Quantifizierung der kanonischen Nukleoside mit einem UV-Detektor. ULOQ: *upper limit of quantification*, LLOQ, *lower limit of quantification*. y = UV Signalfläche A, x = n(A); A: Analyt. **[133, 188]**

Verbindung	*Geradengleichung*	*n (ULOQ)*	*n (LLOQ)*
dC	y = 0.28558x – 0.02812	7751 pmol	10.63 pmol
dG	y = 0.54214x + 0.1554	4892 pmol	19.11 pmol

Tabelle 5.8: Substanzabhängigen LC-MS/MS Geradengleichungen der Quantifizierungskalibriergeraden und die enpsrechenden Gültigkeitsbereiche der Kalibriergerade für die Quantifizierung der markierten Nukleoside mittels Massenspektrometrie. ULOQ: *upper limit of quantification*, LLOQ, *lower limit of quantification*. y = Signal A/Signal A*, x = n(A)/n(A*); A: Analyt, A*: interner Standard. **[133, 188]**

Verbindung	*Geradengleichung*	*n (ULOQ)*	*n (LLOQ)*
dC	$y = 0.92228x + 6.05 \cdot 10^{-3}$	10.6 pmol	20.7 fmol
mdC	$y = 0.751570x – 3.706 \cdot 10^{-4}$	73.45 pmol	100.8 fmol
hmdC	y = 0-89129x – 0.00396	515.0 fmol	2.1 fmol
fdC	y = 1.18055x – 0.00206	496.6 fmol	2.0 fmol
fdC	$y = 0.85582x – 1.22 \cdot 10^{-3}$	20.7 pmol	40.4 fmol
cadC	0.94588x – 0.01218	475.7 fmol	2.0 fmol
fdU	y = 1.29532x – 0.01493	459.6 fmol	5.7 fmol
dT	$y = 0.99777x + 8.11086 \cdot 10^{-4}$	154 pmol	302 fmol

Tabelle 5.9: Substanzabhängigen LC-MS/MS Geradengleichungen der Quantifizierungskalibriergeraden und die enpsrechenden Gültigkeitsbereiche der Kalibriergerade für die Quantifizierung der 2'-fluoerierten Nukleoside mittels Massenspektrometrie. ULOQ: *upper limit of quantification*, LLOQ, *lower limit of quantification*. y = Signal A/Signal A*, x = n(A)/n(A*); A: Analyt, A*: interner Standard. [135, 177, 190]

Verbindung	*Geradengleichung*	*n (ULOQ)*	*n (LLOQ)*
F-fdC	y = 0.8682x + 0.000532	25.1 fmol	0.390 fmol
F-dC	y = 1.0178x + 0.00296	800 fmol	3.13 fmol
F-mdC	y = 0.8693x + 0.000003	49.9 fmol	0.780 fmol
F-cadC	y = 0.8696x + 0.0214	12.7 fmol	0.2 fmol
F-cadC	0.9751x – 0.0203	202.5 fmol	6.3 fmol

Auswertung

Die Auswertung der Daten erfolgte mit *Agilent MassHunter Quantitative Analysis for QQQ* (*Agilent*). Die Darstellung der Daten erfolgte mit Graphpad Prism und Adobe Illustrator.

6 ABKÜRZUNGSVERZEICHNIS

-/- Knockout beider Allele eines Enzyms

+/- Knockout eines der beiden Allele eines Enzyms

° Grad

A Adenosin

A MS-Signal in einer LC-MS/MS des Analyten

A* MS-Signal in einer LC-MS/MS des zugesetzten internen Standards

a-fdC 6-Aza-5-formyl-2'-desoxycytidin

AH Brønsted Säure

AIBN Azobis(isobutyronitril)

Aid Activation induced cytidine deaminase

Ape1 *apurinic/apyrimidinic endonuclease*

Apobec Apolipoprotein B m-RNA editing enzyme catalytic polypeptide

aq. *aqueous*, wässrig

Äq. Äquivalent

Asn Asparagin

ATR *attenuated total reflection*

BER *base excision repair*, Basenexzisionsreparatur

b breit (Signal in einem IR-Spektrum)

B Brønsted Base

BH protonierte Brønsted Base

BHT 2,6-Di-*tert*-butyl-4-methylphenol

C	Cytidin
cadC	5-Carboxyl-2'-desoxycytidin
cadU	5-Carboxyl-2'-desoxyuridin
CAN	Ammoniumcer(IV)-nitrat
CAV	collision cell accelerator voltage
caU	5-Carboxyl-Uridin
CDA	Cytidin Deaminase
CE	*collision energy*, Kollisionsenergie
CMPK1	Cytidin Monophosphat Kinase 1
CpG	dC-dG Dinukleotid
Cys	Cystein
d	Dublet (NMR Spektrum)
dA	2'-Desoxyadenosin
dC	2'-Desoxycytidin
DCM	Dichlormethan
dG	2'-Desoxyguanosin
DKO	*double knockout*
DIPEA	Diisopropylethylamin
DMAP	*N,N*-4-Dimethylaminopyridin
DMF	*N,N*-Dimethylformamid
DMSO	Dimethylsulfoxid
dN	2'-Desoxynukleosid
DNA	*deoxyribonucleic acid*, Desoxyribonukleonsäure

Dnmt	DNA-Methyltransferase
dT	2'-Desoxythymidin
EI	*electron ionization,* Elektronstoß-Ionisation
ESI	*electrospray ionization,* Elektronenspray-Ionisation
EtOAc	Ethylacetat
fdC	5-Formyl-2'-desoxycytidin
fdU	5-Formyl-2'-desoxycytidin
fU	5-Formyl-Uridin
G	Guanosin
gal	Galactosyl
gDNA	genomische DNA
His	Histidin
hmdC	5-Hydroxymethyl-2'-desoxycytidin
hmdU	5-Hydroxymethyl-2'-desoxyuridin
hmU	5-Hydroxymethyl -Uridin
HPLC	*High-performance liquid chromatography*, Hochleistungs-Flüssigchromatographie
HR	*high resolution*, hochaufgelöst
***i*Pr**	Isopropyl
IDcase	Isoorotat Decarboxylase
in vacuo	unter vermindertem Druck
iNGNs	small molecule-inducible neurogenin iPS-line
IR	Infrarotspektroskopie

KG Ketoglutarat

KO *Knockout*

KO *Knockout*

LC *liquid chromatography*

LLOQ *lower limit of quantification,* untere Grenze der Quantifizierung

LOD *limit of detection,* Grenze der Detektion

m mittel (Signal in einem IR-Spektrum)

m Methyl, methyliert

m Multiplet (NMR Spektrum)

MALDI-TOF *Matrix-assisted laser desorption time off light* Massenspektrometrie

man Mannosyl

Mbd *Methyl-CpG-binding domain*

***m*-CPBA** *meta*-Chlorperbenzoesäure

mdC 5-Methyl-2'-desoxycytidin

MeCN Acetonitril

MeOH Methanol

mESCs *murine embryonic stem cells*

MBD *Methhyl-CpG binding domain protein*

MRM *multiple reaction monitoring*

mRNA *messenger ribonucleic acid,* Boten-Ribonukleinsäure

MS Massenspektrometrie / Massenspektrometer

MS/MS Tandemmassenspektrometrie

n Stoffmenge

NCP	*nucleosome core particle*
NeiI	*Nei endonuclease VIII-like*
NEt_3	Triethylamin
NMR	*nuclear magnetic resonance*
nt	Nukleotid
P	Phosphat
PBS Puffer	*Phosphate buffered saline buffer*, Phosphat-gepufferte Salzlösung
ppm	*parts per million*
***p*-TSA**	*para*-Toluolsulfonsäure
q	Quartet (NMR Spektrum)
Q	Queuosin
quant.	quantitativer Umsatz
RNA	*ribonucleic acid,* Ribonukleinsäure
rt	Raumtemperatur
s	Singulet (NMR Spektrum)
s	stark (Signal in einem IR-Spektrum)
SAM	*S*-Adenosylmethionin
SAhC	*S*-Adenosylhomocystein
Smug	*Single-strand selective monofunctional uracil DNA glycosylase*
t	Triplet (NMR Spektrum)
T	Thymin
T7H	Thymin-7-Hydroxylase
TBAPP*i*	Tris(tetrabutylammonium) Pyrophosphat

TBSCl Tertbutyl-dimethylsilylchlorid

Tdg Thymin DNA Glykosylase

Tet *ten eleven translocation*

TFA *trifluoroacetic acid*, Trifluoressigsäure,

TFAA *trifluoroacetic acid anhydride*, Trifluoressigsäure Anhydrid,

TMS Trimethylsilyl

TMSOTf Trifluormethansulfonsäure-trimethylsilylester

TPS-Cl 2,4,6-Triisopropylbenzolsulfonylchlorid

THF Tetrahydrofuran

U Uridin

UHPLC *Ultra high performance liquid chromatography,* Ultrahochleistungs-Flüssigchromatographie

UHRF1 *Ubiquitin-like, containing PHD and RING finger domains, 1,* Zellkernprotein zuständig für Regulation von epigenetischen Modifikationen

ULOD *upper limit of detection*, obere Grenze der Detektion

ULOQ *upper limit of quantification,* obere Grenze der Quantifizierung

vs sehr stark (Signal in einem IR-Spektrum)

w schwach (Signal in einem IR-Spektrum)

wt Wildtyp einer Zelllinie / eines Gens

Ψ Pseudouridin

7 LITERATURVERZEICHNIS

[1] M. R. Mann, M. S. Bartolomei, *Genome Biol.* **2002**, *3*, 1-4.
[2] M. A. Surani, *Nature* **2001**, *414*, 122-128.
[3] J. P. Evans, H. M. Florman, *Nat. Cell. Biol.* **2002**, *4*, 57-63.
[4] R. Jaenisch, R. Young, *Cell* **2008**, *132*, 567-582.
[5] M. J. Evans, M. H. Kaufman, *Nature* **1981**, *292*, 154-156.
[6] J. A. Thomson, J. Itskovitz-Eldor, S. S. Shapiro, M. A. Waknitz, J. J. Swiergiel, V. S. Marshall, J. M. Jones, *Science* **1998**, *282*, 1145-1147.
[7] G. R. Martin, *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 7634-7638.
[8] S. C. Wu, Y. Zhang, *Nat. Rev. Mol. Cell. Biol.* **2010**, *11*, 607-620.
[9] Q. A. Abdul, B. P. Yu, H. Y. Chung, H. A. Jung, J. S. Choi, *Arch. Pharm. Res.* **2017**, *40*, 1219-1237.
[10] C. H. Waddington, *The strategy of the genes*, Routledge, **2014**.
[11] P. A. Jones, S. B. Baylin, *Cell* **2007**, *128*, 683-692.
[12] K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, T. J. Richmond, *Nature* **1997**, *389*, 251-260.
[13] R. D. Kornberg, *Science* **1974**, *184*, 868-871.
[14] F. Song, P. Chen, D. Sun, M. Wang, L. Dong, D. Liang, R.-M. Xu, P. Zhu, G. Li, *Science* **2014**, *344*, 376-380.
[15] G. Felsenfeld, M. Groudine, *Nature* **2003**, *421*, 448-453.
[16] J.-S. Lee, E. Smith, A. Shilatifard, *Cell* **2010**, *142*, 682-685.
[17] T. Jenuwein, C. D. Allis, *Science* **2001**, *293*, 1074-1080.
[18] A. J. Bannister, T. Kouzarides, *Cell Res.* **2011**, *21*, 381-395.
[19] K. E. Gardner, C. D. Allis, B. D. Strahl, *J. Mol. Biol.* **2011**, *409*, 36-46.
[20] H. Wu, Y. Zhang, *Cell* **2014**, *156*, 45-68.
[21] C. Jiang, B. F. Pugh, *Nat. Rev. Genet.* **2009**, *10*, 161-172.
[22] H. Huang, S. Lin, B. A. Garcia, Y. Zhao, *Chem. Rev.* **2015**, *115*, 2376-2418.
[23] J. Wu, M. Grunstein, *Trends Biochem. Sci.* **2000**, *25*, 619-623.
[24] A. F. Kebede, R. Schneider, S. Daujat, *FASEB J.* **2015**, *282*, 1658-1674.
[25] H. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, J. Darnell, in *Molecular Cell Biology. 4th edition*, WH Freeman, **2000**.
[26] T. Kouzarides, *Cell* **2007**, *128*, 693-705.
[27] S. Kumar, V. Chinnusamy, T. Mohapatra, *Front. Genet.* **2018**, *9*.
[28] F. R. Traube, T. Carell, *RNA Biol.* **2017**, *14*, 1099-1107.
[29] A. Sadakierska-Chudy, R. M. Kostrzewa, M. Filip, *Neurotox. Res.* **2015**, *27*, 84-97.
[30] H. Meng, Y. Cao, J. Qin, X. Song, Q. Zhang, Y. Shi, L. Cao, *Int. J. Biol. Sci.* **2015**, *11*, 604-617.
[31] Z. D. Smith, A. Meissner, *Nat. Rev. Genet.* **2013**, *14*, 204-220.
[32] R. Jaenisch, A. Bird, *Nat. Genet.* **2003**, *33*, 245-254.
[33] A. Breiling, F. Lyko, *Epigenetics Chromatin* **2015**, *8*, 24.
[34] R. J. Klose, A. P. Bird, *Trends Biochem. Sci.* **2006**, *31*, 89-97.
[35] T. Pfaffeneder, B. Hackner, M. Truß, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 7008-7012.
[36] R. Z. Jurkowska, T. P. Jurkowski, A. Jeltsch, *Chembiochem* **2011**, *12*, 206-222.
[37] S. Saxonov, P. Berg, D. L. Brutlag, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 1412-1417.
[38] E. Li, *Nat. Rev. Genet.* **2002**, *3*, 662-673.
[39] P. A. Jones, *Trends Genet.* **1999**, *15*, 34-37.
[40] P. A. Jones, D. Takai, *Science* **2001**, *293*, 1068-1070.
[41] P. A. Jones, *Nat. Rev. Genet.* **2012**, *13*, 484-492.

[42] M. L. Gonzalgo, T. Hayashida, C. M. Bender, M. M. Pao, Y. C. Tsai, F. A. Gonzales, H. D. Nguyen, T. T. Nguyen, P. A. Jones, *Cancer Res.* **1998**, *58*, 1245-1252.
[43] X. Nan, H.-H. Ng, C. A. Johnson, C. D. Laherty, B. M. Turner, R. N. Eisenman, A. Bird, *Nature* **1998**, *393*, 386-389.
[44] P. L. Jones, G. J. C. Veenstra, P. A. Wade, D. Vermaak, S. U. Kass, N. Landsberger, J. Strouboulis, A. P. Wolffe, *Nat. Genet.* **1998**, *19*, 187-191.
[45] T. Takizawa, K. Nakashima, M. Namihira, W. Ochiai, A. Uemura, M. Yanagisawa, N. Fujita, M. Nakao, T. Taga, *Dev. Cell* **2001**, *1*, 749-758.
[46] É. Héberlé, Anaïs, *Essays in Biochemistry* **2019**, *63*, 727-741.
[47] M. Okano, D. W. Bell, D. A. Haber, E. Li, *Cell* **1999**, *99*, 247-257.
[48] F. Lyko, B. H. Ramsahoye, H. Kashevsky, M. Tudor, M.-A. Mastrangelo, T. L. Orr-Weaver, R. Jaenisch, *Nat. Genet.* **1999**, *23*, 363-366.
[49] C.-L. Hsieh, *Mol. Cell. Biol.* **1999**, *19*, 8211-8218.
[50] T. H. Bestor, *EMBO Rep.* **1992**, *11*, 2611-2617.
[51] L. Chen, A. M. MacMillan, W. Chang, K. Ezaz-Nikpay, W. S. Lane, G. L. Verdine, *Biochemistry* **1991**, *30*, 11018-11025.
[52] J. C. Wu, D. Santi, *J. Biol. Chem.* **1987**, *262*, 4778-4786.
[53] M. Oka, A. M. Meacham, T. Hamazaki, N. Rodić, L.-J. Chang, N. Terada, *Oncogene* **2005**, *24*, 3091-3099.
[54] G. Vilkaitis, S. Serva, E. Weinhold, S. Klimašauskas, *J. Biol. Chem.* **2001**, *276*, 20924-20934.
[55] A. Smith, *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 4744-4748.
[56] X. Lu, B. S. Zhao, C. He, *Chem. Rev.* **2015**, *115*, 2225-2239.
[57] C.-W. J. Lio, X. Yue, I. F. López-Moyado, M. Tahiliani, L. Aravind, A. Rao, *J. Biosci.* **2020**, *45*.
[58] H. Wu, Y. Zhang, *Genes Dev.* **2011**, *25*, 2436-2452.
[59] T. Carell, M. Q. Kurz, M. Müller, M. Rossa, F. Spada, *Angew. Chem. Int. Ed.* **2018**, *57*, 4296-4312.
[60] D. Svilar, E. M. Goellner, K. H. Almeida, R. W. Sobol, *Antioxid. Redox Signal.* **2011**, *14*, 2491-2507.
[61] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929-930.
[62] M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930-935.
[63] D. Globisch, M. Münzel, M. Müller, S. Michalakis, M. Wagner, S. Koch, T. Bruckl, M. Biel, T. Carell, *PLoS One* **2010**, *5*, e15367.
[64] M. Münzel, D. Globisch, T. Bruckl, M. Wagner, V. Welzmiller, S. Michalakis, M. Müller, M. Biel, T. Carell, *Angew. Chem. Int. Ed.* **2010**, *49*, 5375-5377.
[65] M. Münzel, D. Globisch, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 6460-6468.
[66] M. Wagner, J. Steinbacher, T. F. Kraus, S. Michalakis, B. Hackner, T. Pfaffeneder, A. Perera, M. Müller, A. Giese, H. A. Kretzschmar, *Angew. Chem. Int. Ed.* **2015**, *54*, 12511-12514.
[67] S. Dzitoyeva, H. Chen, H. Manev, *Neurobiol. Aging* **2012**, *33*, 2881-2891.
[68] W. Sun, L. Zang, Q. Shu, X. Li, *Genomics* **2014**, *104*, 347-351.
[69] T. Pfaffeneder, B. Hackner, M. Truss, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 7008-7012.
[70] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300-1303.
[71] Y. F. He, B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He, G. L. Xu, *Science* **2011**, *333*, 1303-1307.
[72] S. Liu, J. Wang, Y. Su, C. Guerrero, Y. Zeng, D. Mitra, P. J. Brooks, D. E. Fisher, H. Song, Y. Wang, *Nucleic Acids Res.* **2013**, *41*, 6421-6429.
[73] S. Kal, L. Que, *J. Biol. Inorg. Chem.* **2017**, *22*, 339-365.

[74] R. P. Hausinger, in *2-Oxoglutarate-Dependent Oxygenases* Royal Society of Chemistry, Cambridge, **2015**.
[75] L. Hu, Z. Li, J. Cheng, Q. Rao, W. Gong, M. Liu, Y. G. Shi, J. Zhu, P. Wang, Y. Xu, *Cell* **2013**, *155*, 1545-1555.
[76] H. Hashimoto, J. E. Pais, X. Zhang, L. Saleh, Z.-Q. Fu, N. Dai, I. R. Corrêa, Y. Zheng, X. Cheng, *Nature* **2014**, *506*, 391-395.
[77] H. Hashimoto, J. E. Pais, N. Dai, I. R. Corrêa Jr, X. Zhang, Y. Zheng, X. Cheng, *Nucleic Acids Res.* **2015**, *43*, 10713-10721.
[78] F. P. Guengerich, *J. Biol. Chem.* **2015**, *290*, 20700-20701.
[79] Y. Mishina, C. He, *J. Inorg. Biochem.* **2006**, *100*, 670-678.
[80] Y. Mishina, L. X. Chen, C. He, *J. Am. Chem. Soc.* **2004**, *126*, 16930-16936.
[81] H. Liu, J. Llano, J. W. Gauld, *J. Phys. Chem.* **2009**, *113*, 4887-4898.
[82] L. Que Jr, *Acc. Chem. Res.* **2007**, *40*, 493-500.
[83] B. Bleijlevens, T. Shivarattan, E. Flashman, Y. Yang, P. J. Simpson, P. Koivisto, B. Sedgwick, C. J. Schofield, S. J. Matthews, *EMBO Rep.* **2008**, *9*, 872-877.
[84] B. Yu, W. C. Edstrom, J. Benach, Y. Hamuro, P. C. Weber, B. R. Gibney, J. F. Hunt, *Nature* **2006**, *439*, 879-884.
[85] W. Mayer, A. Niveleau, J. Walter, R. Fundele, T. Haaf, *Nature* **2000**, *403*, 501-502.
[86] M. Wossidlo, T. Nakamura, K. Lepikhov, C. J. Marques, V. Zakhartchenko, M. Boiani, J. Arand, T. Nakano, W. Reik, J. Walter, *Nat. Commun.* **2011**, *2*, 1-8.
[87] M. Bachman, S. Uribe-Lewis, X. Yang, H. E. Burgess, M. Iurlaro, W. Reik, A. Murrell, S. Balasubramanian, *Nat. Chem. Biol.* **2015**, *11*, 555-557.
[88] M. Su, A. Kirchner, S. Stazzoni, M. Müller, M. Wagner, A. Schröder, T. Carell, *Angew. Chem. Int. Ed.* **2016**, *55*, 11797-11800.
[89] S. K. Bhattacharya, S. Ramchandani, N. Cervoni, M. Szyf, *Nature* **1999**, *397*, 579-583.
[90] H. Cedar, G. L. Verdine, *Nature* **1999**, *397*, 568-569.
[91] B. Hendrich, J. Guy, B. Ramsahoye, V. A. Wilson, A. Bird, *Genes Dev.* **2001**, *15*, 710-723.
[92] F. Santos, B. Hendrich, W. Reik, W. Dean, *Dev. Biol.* **2002**, *241*, 172-182.
[93] S. Schiesser, B. Hackner, T. Pfaffeneder, M. Müller, C. Hagemeier, M. Truss, T. Carell, *Angew. Chem. Int. Ed.* **2012**, *51*, 6516-6520.
[94] S. Cortellino, J. Xu, M. Sannai, R. Moore, E. Caretti, A. Cigliano, M. Le Coz, K. Devarajan, A. Wessels, D. Soprano, L. K. Abramowitz, M. S. Bartolomei, F. Rambow, M. R. Bassi, T. Bruno, M. Fanciulli, C. Renner, A. J. Klein-Szanto, Y. Matsumoto, D. Kobi, I. Davidson, C. Alberti, L. Larue, A. Bellacosa, *Cell* **2011**, *146*, 67-79.
[95] A. Maiti, A. C. Drohat, *J. Biol. Chem.* **2011**, *286*, 35334-35338.
[96] A. Maiti, A. Z. Michelson, C. J. Armwood, J. K. Lee, A. C. Drohat, *J. Am. Chem. Soc.* **2013**, *135*, 15813-15822.
[97] J. C. Fromme, A. Banerjee, G. L. Verdine, *Curr. Opin. Struct. Biol.* **2004**, *14*, 43-49.
[98] L. Serre, K. P. de Jésus, S. Boiteux, C. Zelwer, B. Castaing, *EMBO Rep.* **2002**, *21*, 2854-2865.
[99] A. C. Drohat, A. Maiti, *Org. Biomol. Chem.* **2014**, *12*, 8367-8378.
[100] P. Fortini, E. Dogliotti, *DNA Repair* **2007**, *6*, 398-409.
[101] A. J. Podlutsky, I. I. Dianova, S. H. Wilson, V. A. Bohr, G. L. Dianov, *Biochemistry* **2001**, *40*, 809-813.
[102] A. C. Drohat, C. T. Coey, *Chem. Rev.* **2016**, *116*, 12711-12729.
[103] R. Métivier, R. Gallais, C. Tiffoche, C. Le Peron, R. Z. Jurkowska, R. P. Carmouche, D. Ibberson, P. Barath, F. Demay, G. Reid, V. Benes, A. Jeltsch, F. Gannon, G. Salbert, *Nature* **2008**, *452*, 45-50.
[104] H. D. Morgan, W. Dean, H. A. Coker, W. Reik, S. K. Petersen-Mahrt, *J. Biol. Chem.* **2004**, *279*, 52353-52360.
[105] E. K. Schutsky, C. S. Nabel, A. K. F. Davis, J. E. DeNizio, R. M. Kohli, *Nucleic Acids Res.* **2017**, *45*, 7655-7665.

[106] C. Popp, W. Dean, S. H. Feng, S. J. Cokus, S. Andrews, M. Pellegrini, S. E. Jacobsen, W. Reik, *Nature* **2010**, *463*, 1101-1126.
[107] N. Bhutani, J. J. Brady, M. Damian, A. Sacco, S. Y. Corbel, H. M. Blau, *Nature* **2010**, *463*, 1042-1047.
[108] N. Bhutani, M. N. Decker, J. J. Brady, R. T. Bussat, D. M. Burns, S. Y. Corbel, H. M. Blau, *FASEB J.* **2013**, *27*, 1107-1113.
[109] R. Kumar, L. DiMenna, N. Schrode, T.-C. Liu, P. Franck, S. Muñoz-Descalzo, A.-K. Hadjantonakis, A. A. Zarrin, J. Chaudhuri, O. Elemento, *Nature* **2013**, *500*, 89-92.
[110] P. Revy, T. Muto, Y. Levy, F. Geissmann, A. Plebani, O. Sanal, N. Catalan, M. Forveille, R. Dufourcq-Lagelouse, A. Gennery, *Cell* **2000**, *102*, 565-575.
[111] M. Muramatsu, K. Kinoshita, S. Fagarasan, S. Yamada, Y. Shinkai, T. Honjo, *Cell* **2000**, *102*, 553-563.
[112] C. S. Nabel, H. J. Jia, Y. Ye, L. Shen, H. L. Goldschmidt, J. T. Stivers, Y. Zhang, R. M. Kohli, *Nat. Chem. Biol.* **2012**, *8*, 751-758.
[113] K. A. Haushalter, P. T. Stukenberg, M. W. Kirschner, G. L. Verdine, *Curr. Biol.* **1999**, *9*, 174-185.
[114] H. Hashimoto, Y. Liu, A. K. Upadhyay, Y. Chang, S. B. Howerton, P. M. Vertino, X. Zhang, X. Cheng, *Nucleic Acids Res.* **2012**, *40*, 4841-4849.
[115] T. Pfaffeneder, F. Spada, M. Wagner, C. Brandmayr, S. K. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarova, D. Schürmann, S. Michalakis, O. Kosmatchev, S. Schiesser, B. Steigenberger, N. Raddaoui, G. Kashiwazaki, U. Müller, C. G. Spruijt, M. Vermeulen, H. Leonhardt, P. Schär, M. Müller, T. Carell, *Nat. Chem. Biol.* **2014**, *10*, 574-581.
[116] R. J. Boorstein, A. Cummings, D. R. Marenstein, M. K. Chan, Y. L. Ma, T. A. Neubert, S. M. Brown, G. W. Teebor, *J. Biol. Chem.* **2001**, *276*, 41991-41997.
[117] K. L. Limpose, K. S. Trego, Z. Li, S. W. Leung, A. H. Sarker, J. A. Shah, S. S. Ramalingam, E. M. Werner, W. S. Dynan, P. K. Cooper, *Nucleic Acids Res.* **2018**, *46*, 4515-4532.
[118] P. L. Olive, *Radiat. Res.* **1998**, *150*, 42-51.
[119] J. H. Hoeijmakers, *Nature* **2001**, *411*, 366-374.
[120] C. X. Song, K. E. Szulwach, Q. Dai, Y. Fu, S. Q. Mao, L. Lin, C. Street, Y. Li, M. Poidevin, H. Wu, J. Gao, P. Liu, L. Li, G. L. Xu, P. Jin, C. He, *Cell* **2013**, *153*, 678-691.
[121] L. Shen, H. Wu, D. Diep, S. Yamaguchi, A. C. D'Alessio, H. L. Fung, K. Zhang, Y. Zhang, *Cell* **2013**, *153*, 692-706.
[122] F. Neri, D. Incarnato, A. Krepelova, S. Rapelli, F. Anselmi, C. Parlato, C. Medana, F. Dal Bello, S. Oliviero, *Cell Rep.* **2015**, *10*, 674-683.
[123] R. Rahimoff, O. Kosmatchev, A. Kirchner, T. Pfaffeneder, F. Spada, V. Brantl, M. Müller, T. Carell, *J. Am. Chem. Soc.* **2017**, *139*, 10359-10364.
[124] A. R. Weber, C. Krawczyk, A. B. Robertson, A. Kusnierczyk, C. B. Vagbo, D. Schuermann, A. Klungland, P. Schar, *Nat. Commun.* **2016**, *7*, 10806.
[125] L. Schomacher, D. D. Han, M. U. Musheev, K. Arab, S. Kienhofer, A. von Seggern, C. Niehrs, *Nat. Struct. Mol. Biol.* **2016**, *23*, 116-124.
[126] A. L. Jacobs, P. Schär, *Chromosoma* **2012**, *121*, 1-20.
[127] I. Morland, V. Rolseth, L. Luna, T. Rognes, M. Bjørås, E. Seeberg, *Nucleic Acids Res.* **2002**, *30*, 4926-4936.
[128] Z. Liutkevičiūtė, G. Lukinavičius, V. Masevičius, D. Daujotytė, S. Klimašauskas, *Nat. Chem. Biol.* **2009**, *5*, 400-402.
[129] C. G. Spruijt, F. Gnerlich, A. H. Smits, T. Pfaffeneder, P. W. Jansen, C. Bauer, M. Münzel, M. Wagner, M. Müller, F. Khan, H. C. Eberl, A. Mensinga, A. B. Brinkman, K. Lephikov, U. Müller, J. Walter, R. Boelens, H. van Ingen, H. Leonhardt, T. Carell, M. Vermeulen, *Cell* **2013**, *152*, 1146-1159.
[130] J. C. Fromme, G. L. Verdine, in *DNA Repair and Replication, Vol. 69*, **2004**, pp. 1-41.

[131] F. Guo, X. Li, D. Liang, T. Li, P. Zhu, H. Guo, X. Wu, L. Wen, T.-P. Gu, B. Hu, Colum P. Walsh, J. Li, F. Tang, G.-L. Xu, *Cell Stem Cell* **2014**, *15*, 447-458.
[132] C. C. Chen, K. Y. Wang, C. K. Shen, *J. Biol. Chem.* **2013**, *288*, 9084-9091.
[133] S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Höfner, K. T. Wanner, C. Ochsenfeld, T. Carell, *J. Am. Chem. Soc.* **2013**, *135*, 14593-14599.
[134] M. Münzel, D. Globisch, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 6460-6468.
[135] K. Iwan, R. Rahimoff, A. Kirchner, F. Spada, A. S. Schröder, O. Kosmatchev, S. Ferizaj, J. Steinbacher, E. Parsa, M. Müller, T. Carell, *Nat. Chem. Biol.* **2018**, *14*, 72-78.
[136] Y. Feng, N.-B. Xie, W.-B. Tao, J.-H. Ding, X.-J. You, C.-J. Ma, X. Zhang, C. Yi, X. Zhou, B.-F. Yuan, Y.-Q. Feng, *CCS Chem.* **2021**, *3*, 994-1008.
[137] Z. Liutkeviciute, E. Kriukiene, J. Licyte, M. Rudyte, G. Urbanaviciute, S. Klimasauskas, *J. Am. Chem. Soc.* **2014**, *136*, 5884-5887.
[138] J. A. Smiley, J. M. Angelot, R. C. Cannon, E. M. Marshall, D. K. Asch, *Anal. Biochem.* **1999**, *266*, 85-92.
[139] S. Xu, W. Li, J. Zhu, R. Wang, Z. Li, G.-L. Xu, J. Ding, *Cell Res.* **2013**, *23*, 1296-1309.
[140] M. T. Abbott, E. K. Schandl, R. F. Lee, T. S. Parker, R. J. Midgett, *Biochim. Biophys. Acta* **1967**, *132*, 525-528.
[141] M. S. Watanabe, R. P. McCroskey, M. T. Abbott, *J. Biol. Chem.* **1970**, *245*, 2023-2026.
[142] E. Holme, G. Lindstedt, S. Lindstedt, M. Tofft, *J. Biol. Chem.* **1971**, *246*, 3314-3319.
[143] C. K. Liu, P. M. Shaffer, R. S. Slaughter, R. P. McCroskey, M. T. Abbott, *Biochemistry* **1972**, *11*, 2172-2176.
[144] C.-K. Liu, C.-A. Hsu, M. T. Abbott, *Arch. Biochem. Biophys.* **1973**, *159*, 180-187.
[145] P. Boccaletto, M. A. Machnicka, E. Purta, P. Piątkowski, B. Bagiński, T. K. Wirecki, Valérie, R. Ross, P. A. Limbach, A. Kotter, M. Helm, J. M. Bujnicki, *Nucleic Acids Res.* **2018**, *46*, 303-307.
[146] W. A. Cantara, P. F. Crain, J. Rozenski, J. A. McCloskey, K. A. Harris, X. Zhang, F. A. P. Vendeix, D. Fabris, P. F. Agris, *Nucleic Acids Res.* **2011**, *39*, 195-201.
[147] M. A. Machnicka, K. Milanowska, O. Osman Oglou, E. Purta, M. Kurkowska, A. Olchowik, W. Januszewski, S. Kalinowski, S. Dunin-Horkawicz, K. M. Rother, M. Helm, J. M. Bujnicki, H. Grosjean, *Nucleic Acids Res.* **2012**, *41*, 262-267.
[148] X. Wang, Z. Lu, A. Gomez, G. C. Hon, Y. Yue, D. Han, Y. Fu, M. Parisien, Q. Dai, G. Jia, B. Ren, T. Pan, C. He, *Nature* **2014**, *505*, 117-120.
[149] N. Sonenberg, M. A. Morgan, W. C. Merrick, A. J. Shatkin, *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 4843-4847.
[150] X. Zhao, Y. Yang, B.-F. Sun, Y. Shi, X. Yang, W. Xiao, Y.-J. Hao, X.-L. Ping, Y.-S. Chen, W.-J. Wang, K.-X. Jin, X. Wang, C.-M. Huang, Y. Fu, X.-M. Ge, S.-H. Song, H. S. Jeong, H. Yanagisawa, Y. Niu, G.-F. Jia, W. Wu, W.-M. Tong, A. Okamoto, C. He, J. M. R. Danielsen, X.-J. Wang, Y.-G. Yang, *Cell Res.* **2014**, *24*, 1403-1419.
[151] W. Xiao, S. Adhikari, U. Dahal, Y.-S. Chen, Y.-J. Hao, B.-F. Sun, H.-Y. Sun, A. Li, X.-L. Ping, W.-Y. Lai, X. Wang, H.-L. Ma, C.-M. Huang, Y. Yang, N. Huang, G.-B. Jiang, H.-L. Wang, Q. Zhou, X.-J. Wang, Y.-L. Zhao, Y.-G. Yang, *Mol. Cell* **2016**, *61*, 507-519.
[152] P. C. Thiaville, B. El Yacoubi, C. Köhrer, J. J. Thiaville, C. Deutsch, D. Iwata-Reuyl, J. M. Bacusmo, J. Armengaud, Y. Bessho, C. Wetzel, X. Cao, P. A. Limbach, U. L. Rajbhandary, V. De Crécy-Lagard, *Mol. Microbiol.* **2015**, *98*, 1199-1221.
[153] J. Zhou, J. Wan, X. Gao, X. Zhang, S. R. Jaffrey, S.-B. Qian, *Nature* **2015**, *526*, 591-594.
[154] D. P. Patil, C.-K. Chen, B. F. Pickering, A. Chow, C. Jackson, M. Guttman, S. R. Jaffrey, *Nature* **2016**, *537*, 369-373.
[155] W. E. Cohn, *J. Biol. Chem.* **1960**, *235*, 1488-1498.

[156] S. Yokoyama, T. Miyazawa, Y. Iitaka, Z. Yamaizumi, H. Kasai, S. Nishimura, *Nature* **1979**, *282*, 107-109.
[157] H. Kasai, K. Nakanishi, R. D. Macfarlane, D. F. Torgerson, Z. Ohashi, J. A. McCloskey, H. J. Gross, S. Nishimura, *J. Am. Chem. Soc.* **1976**, *98*, 5044-5046.
[158] T. Carell, C. Brandmayr, A. Hienzsch, M. Müller, D. Pearson, V. Reiter, I. Thoma, P. Thumbs, M. Wagner, *Angew. Chem. Int. Ed.* **2012**, *51*, 7110-7131.
[159] F. Harada, S. Nishimura, *Biochemistry* **1972**, *11*, 301-308.
[160] D. W. Phillipson, C. G. Edmonds, P. F. Crain, D. L. Smith, D. R. Davis, J. A. McCloskey, *J. Biol. Chem.* **1987**, *262*, 3462-3471.
[161] J. C. Salazar, A. Ambrogelly, P. F. Crain, J. A. McCloskey, D. Soll, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7536-7541.
[162] M. Blaise, H. D. Becker, J. Lapointe, C. Cambillau, R. Giegé, D. Kern, *Biochimie* **2005**, *87*, 847-861.
[163] U. Gündüz, J. R. Katze, *Biochem. Biophys. Res. Commun.* **1982**, *109*, 159-167.
[164] R. K. Slany, M. Bösl, H. Kersten, *Biochimie* **1994**, *76*, 389-393.
[165] Z. D. Miles, R. M. McCarty, G. Molnar, V. Bandarian, *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 7368-7372.
[166] D. Iwata-Reuyl, *Bioorg. Chem.* **2003**, *31*, 24-43.
[167] R. M. McCarty, C. Krebs, V. Bandarian, *Biochemistry* **2013**, *52*, 188-198.
[168] Z. D. Miles, W. K. Myers, W. M. Kincannon, R. D. Britt, V. Bandarian, *Biochemistry* **2015**, *54*, 4927-4935.
[169] M. A. Alqasem, C. Fergus, J. M. Southern, S. J. Connon, V. P. Kelly, *Chem. Commun.* **2020**, *56*, 3915-3918.
[170] S. Johannsson, P. Neumann, R. Ficner, *Biomolecules* **2018**, *8*, 81.
[171] T. Rakovich, C. Boland, I. Bernstein, V. M. Chikwana, D. Iwata-Reuyl, V. P. Kelly, *J. Biol. Chem.* **2011**, *286*, 19354-19363.
[172] W. Baranowski, G. Dirheimer, J. A. Jakowicki, G. Keith, *Cancer Res.* **1994**, *54*, 4468-4471.
[173] B. Emmerich, E. Zubrod, H. Weber, P. A. Maubach, H. Kersten, W. Kersten, *Cancer Res.* **1985**, *45*, 4308-4314.
[174] B.-S. Huang, R.-T. Wu, K.-Y. Chien, *Cancer Res.* **1992**, *52*, 4696-4700.
[175] P. Thumbs, T. T. Ensfelder, M. Hillmeier, M. Wagner, M. Heiss, C. Scheel, A. Schön, M. Müller, S. Michalakis, S. Kellner, T. Carell, *Angew. Chem. Int. Ed.* **2020**, *59*, 12352-12356.
[176] P. Thumbs, PhD Thesis thesis, LMU Munich **2013**.
[177] E. Korytiaková, E. Kamińska, M. Müller, T. Carell, *Angew. Chem. Int. Ed.* **2021**, *60*, 16869-16873.
[178] Z. Zawada, A. Tatar, P. Mocilac, M. Buděšínský, T. Kraus, *Angew. Chem. Int. Ed.* **2018**, *57*, 9891-9895.
[179] J. C. Price, E. W. Barr, B. Tirupati, J. M. Bollinger, C. Krebs, *Biochemistry* **2003**, *42*, 7497-7508.
[180] N. S. W. Jonasson, L. J. Daumann, *Chemistry* **2019**, *25*, 12091-12097.
[181] P. J. Riggs-Gelasco, J. C. Price, R. B. Guyer, J. H. Brehm, E. W. Barr, J. M. Bollinger, C. Krebs, *J. Am. Chem. Soc.* **2004**, *126*, 8108-8109.
[182] M. L. Neidig, C. D. Brown, K. M. Light, D. G. Fujimori, E. M. Nolan, J. C. Price, E. W. Barr, J. M. Bollinger, C. Krebs, C. T. Walsh, E. I. Solomon, *J. Am. Chem. Soc.* **2007**, *129*, 14224-14231.
[183] A. S. Schröder, E. Parsa, K. Iwan, M. Wallner, S. Serdjukow, T. Carell, *Chem. Commun.* **2016**, *52*, 14361-14364.
[184] A. S. Schröder, O. Kotljarova, E. Parsa, K. Iwan, N. Raddaoui, T. Carell, *Org. Lett.* **2016**, *18*, 4368-4371.
[185] Y. Saito, T. A. Zevaco, L. A. Agrofoglio, *Tetrahedron* **2002**, *58*, 9593-9603.
[186] M. Zauri, G. Berridge, M.-L. Thézénas, K. M. Pugh, R. Goldin, B. M. Kessler, S. Kriaucionis, *Nature* **2015**, *524*, 114-118.

[187] R. J. Boorstein, L. N. Chiu, G. W. Teebor, *Mol. Cell. Biol.* **1992**, *12*, 5536-5540.
[188] T. Pfaffeneder, F. Spada, M. Wagner, C. Brandmayr, S. K. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarova, D. Schuermann, S. Michalakis, O. Kosmatchev, S. Schiesser, B. Steigenberger, N. Raddaoui, G. Kashiwazaki, U. Müller, C. G. Spruijt, M. Vermeulen, H. Leonhardt, P. Schär, M. Müller, T. Carell, *Nat. Chem. Biol.* **2014**, *10*, 574-581.
[189] T. Carell, C. Brandmayr, A. Hienzsch, M. Müller, D. Pearson, V. Reiter, I. Thoma, P. Thumbs, M. Wagner, *Angew. Chem. Int. Ed.* **2012**, *51*, 7110-7131.
[190] E. Kamińska, E. Korytiaková, A. Reichl, M. Müller, T. Carell, *Angew. Chem. Int. Ed.* **2021**, *60*, 23207–23211.
[191] Y. Feng, J.-J. Chen, N.-B. Xie, J.-H. Ding, X.-J. You, W.-B. Tao, X. Zhang, C. Yi, X. Zhou, B.-F. Yuan, Y.-Q. Feng, *Chem. Sci.* **2021**, *12*, 11322–11329.
[192] G. Song, G. Wang, X. Luo, Y. Cheng, Q. Song, J. Wan, C. Moore, H. Song, P. Jin, J. Qian, H. Zhu, *Nat. Commun.* **2021**, *12*, 795.
[193] C. Meier, *Antivir. Chem. Chemother.* **2017**, *25*, 69-82.
[194] T. Gollnest, T. D. de Oliveira, D. Schols, J. Balzarini, C. Meier, *Nat. Commun.* **2015**, *6*, 8716.
[195] T. Gollnest, T. Dinis de Oliveira, A. Rath, I. Hauber, D. Schols, J. Balzarini, C. Meier, *Angew. Chem. Int. Ed.* **2016**, *55*, 5255-5258.
[196] X. Jia, D. Schols, C. Meier, *J. Med. Chem.* **2020**, *63*, 6991-7007.
[197] X. Jia, S. Weber, D. Schols, C. Meier, *J. Med. Chem.* **2020**, *63*, 11990-12007.
[198] S. S. Ibrahim, F. D. Boudinot, R. F. Schinazi, C. K. Chu, *Antivir. Chem. Chemother.* **1996**, *7*, 167-172.
[199] A. Schön, E. Kamińska, F. Schelter, E. Ponkkonen, E. Korytiaková, S. Schiffers, T. Carell, *Angew. Chem. Int. Ed.* **2020**, *59*, 5591-5594.
[200] J.-I. Asakura, M. J. Robins, *Tetrahedron Lett.* **1988**, *29*, 2855-2858.
[201] S. Meneni, I. Ott, C. D. Sergeant, A. Sniady, R. Gust, R. Dembinski, *Bioorg. Med. Chem.* **2007**, *15*, 3082-3088.
[202] M. Aso, T. Kaneko, M. Nakamura, N. Koga, H. Suemune, *Chem. Commun.* **2003**, 1094-1095.
[203] F. R. Traube, S. Schiffers, K. Iwan, S. Kellner, F. Spada, M. Müller, T. Carell, *Nat. Protoc.* **2019**, *14*, 283-312.

8 ANHANG

8.1 Supporting Information: Deformylation of 5-formylcytidine in different cell types

E. Korytiaková*, E. Kaminska*, M. Müller, T. Carell *Angew. Chem. Int. Ed.* **2021**, *60,* 16869-16873.

*: Diese Autoren haben zu gleichen Teilen beigetragen

Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

Deformylation of 5-Formylcytidine in Different Cell Types

Eva Korytiaková[+], *Ewelina Kamińska*[+], *Markus Müller, and Thomas Carell**

anie_202107089_sm_miscellaneous_information.pdf

Table of Contents

Supplementary figures 2
Methods and Materials 10
Chemical Synthesis 10
Cell culture 11
Cell Harvesting and Isolation of genomic DNA 12
DNA digestion 12
LC/MS-MS analysis of DNA samples 13
Quantification of nucleosides in DNA samples 13
Soluble nucleoside pool extraction and purification 13
Synthetic procedures 15
Synthesis of O3′,5′-Bis(tertbutyldimethylsilyl)-2′-fluorouridine (**11**) 16
Synthesis of O3′,5′-Bis(tertbutyldimethylsilyl)-2′-fluor-5-iodouridine (**5**) 18
Synthesis of O3′,5′-Bis(tertbutyldimethylsilyl)-2′-fluor-5-formyluridine (**6**) 20
Synthesis of O3′,5′-Bis(tertbutyldimethylsilyl)-2′-fluor-5-(1,3-dioxan-2-yl)uridine (**7**) 22
Synthesis of O3′,5′-Bis(tertbutyldimethylsilyl)-2′-fluor-5-(1,3-dioxan-2-yl)cytidine (**8**) 24
Synthesis of 2′-Fluor-5-formyl-cytidine (**1**) 26
Supplementary tables 28
Appendix 31
NMR Spectra 31
References 49

Supplementary figures

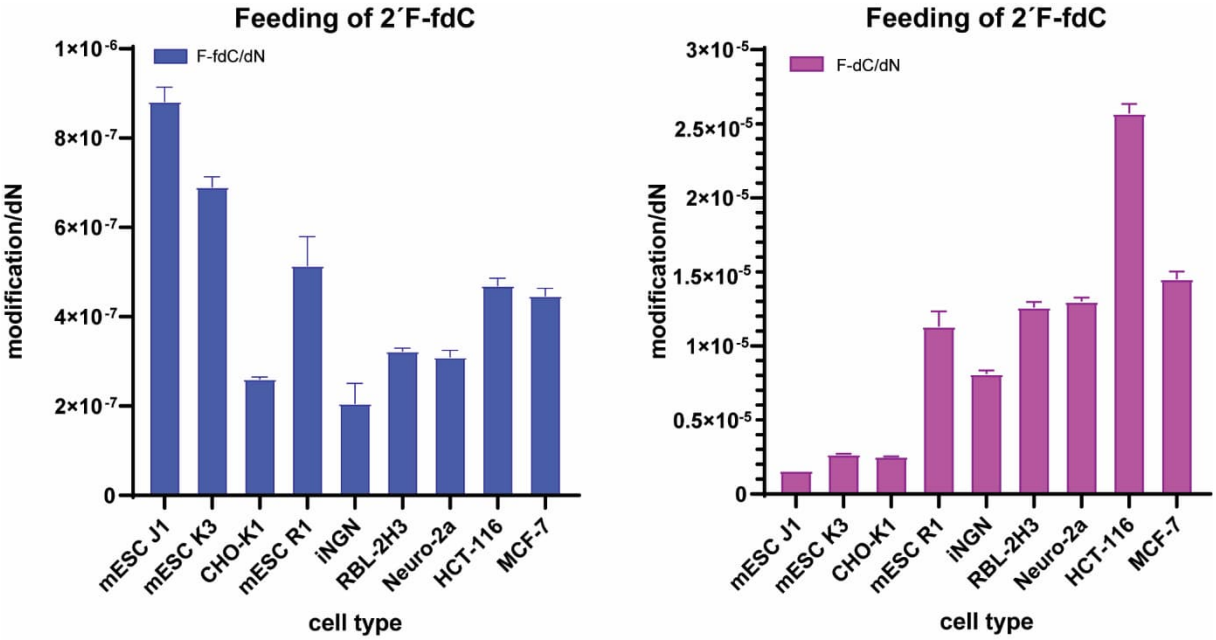


Figure SI-1: Levels of incorporated 2´-F-fdC (**1**) in the genome (left) and of the deformylation product 2´-F-dC (**2**, right) of different cells after feeding 2´F-fdC at a concentration of 350 μM. The error bards represent the propagated error based on the standard deviations of three biological replicates.

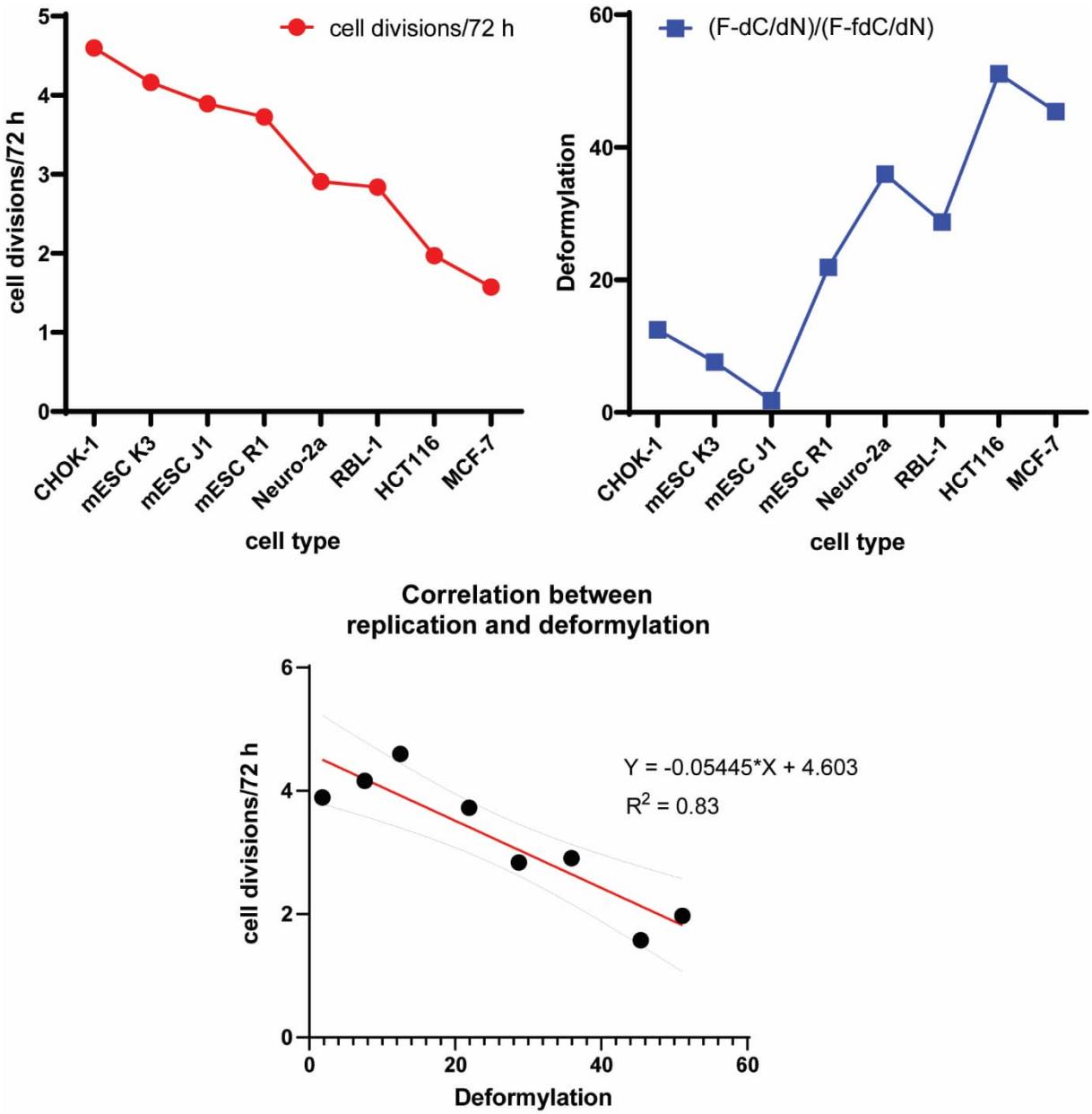


Figure SI-2A: Upper graphs represent the absolute number of the cell divisions calculated as $doubling\ time = \frac{72*\log(2)}{log_2(F)-log_2(I)}$, F = final number of cells, I = initial number of cells. and the rates of deformylation. The lower graph shows the inverse correlation between cell doublings over the duration of the experiment and their corresponding deformylation rates.

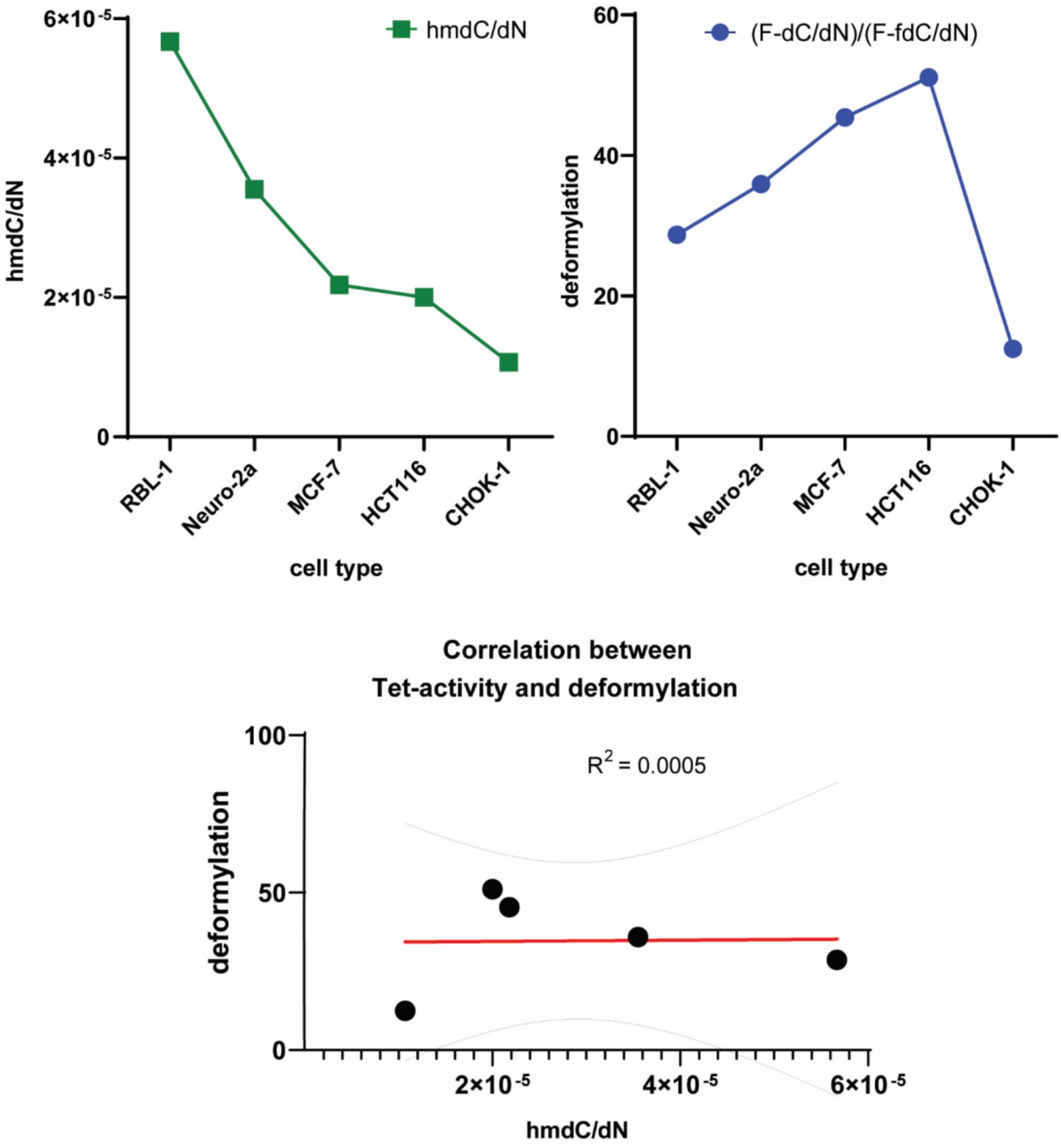


Figure SI-2B: Upper graphs represent the levels of genomic natural hmdC per dN in different cell lines and their corresponding rates of deformylation. The lower graph shows no correlation between Tet – enzymes activity and deformylation as a result of the analysis.

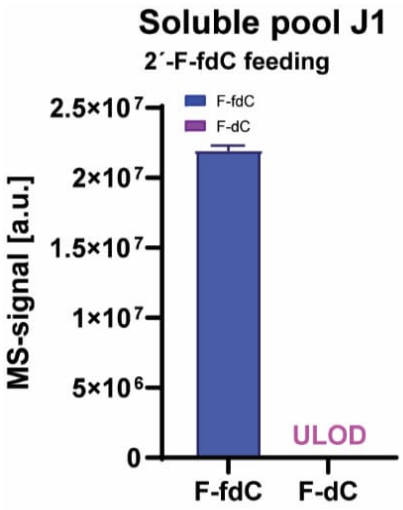


Figure SI-3A: *In vivo* stability assay of 2´F-fdC (**1**). **1** was incubated for 72 h in the cellular medium of J1 at 37 °C. Soluble pool content after feeding 2´F-fdC (**1**) at a concentration of 350 μM, levels of 2´F-fdC (**1**) vs. 2´F-dC (**2**). Error bars represent standard deviation.

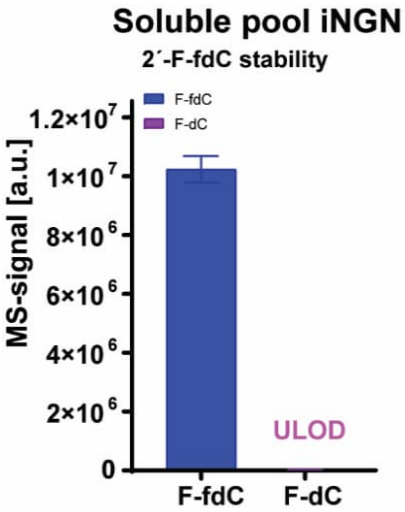


Figure SI-3B: *In vivo* Stability assay of 2´F-fdC (**1**). **1** was incubated for 24 h in the cellular medium of iNGNs at 37 °C. Levels of 2´F-fdC (**1**) vs. 2´F-dC (**2**) analyzed by UHPLC-MS/MS. Error bars represent standard deviation.

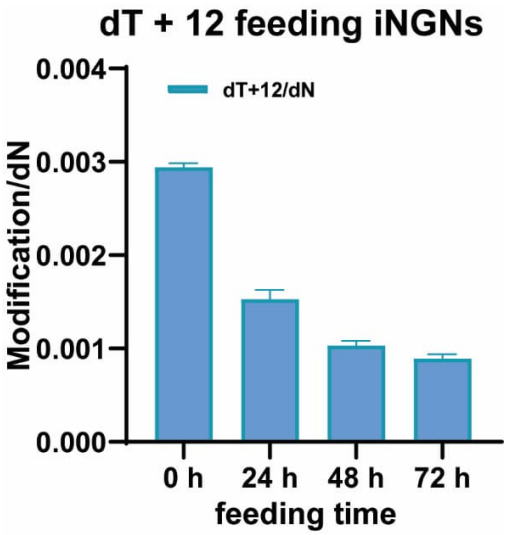


Figure SI-4: Amount of dT+12 (**10**) incorporated in the genome of iNGN cells to different time points after feeding **10**. The error bars represent standard deviation.

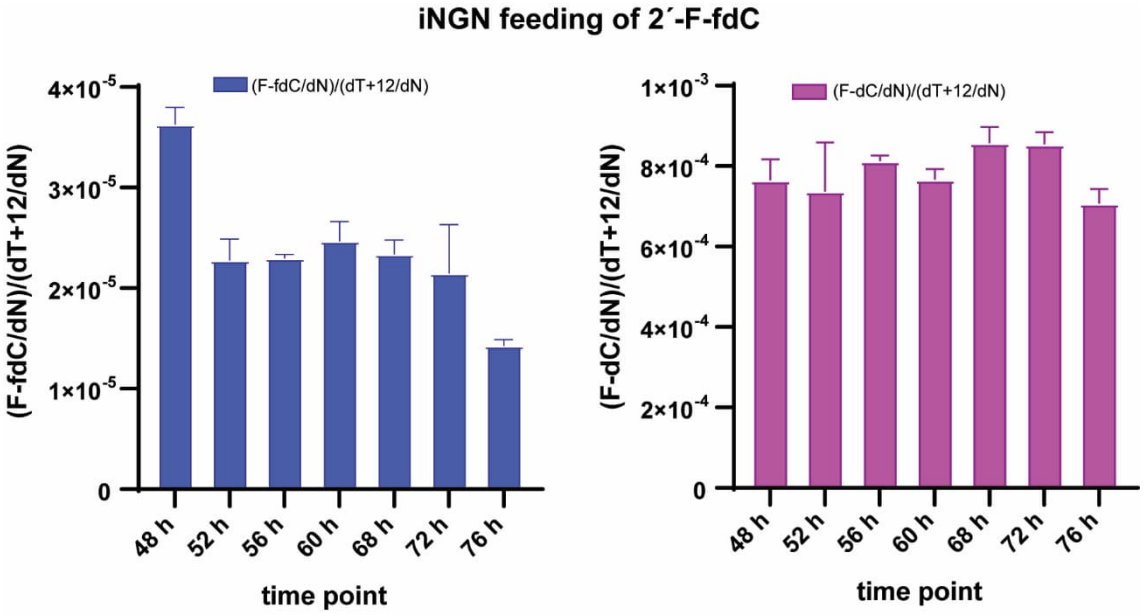


Figure SI-5: Levels of incorporated 2´-F-fdC (**1**) in the genome (left) and of the deformylation product 2´-F-dC (**2**, right) of iNGNs normalized to cell division (dT+12/dN) to different time points after feeding of 2´F-fdC (**1**) at a concentration of 350 µM. While levels of 2´F-fdC/dN decrease, this cannot be directly mirrored by the increase in the levels of 2´F-dC/dN between different time points, due to the fact, that the levels of these modifications are one order of magnitude apart. The error bars represent the standard deviation-

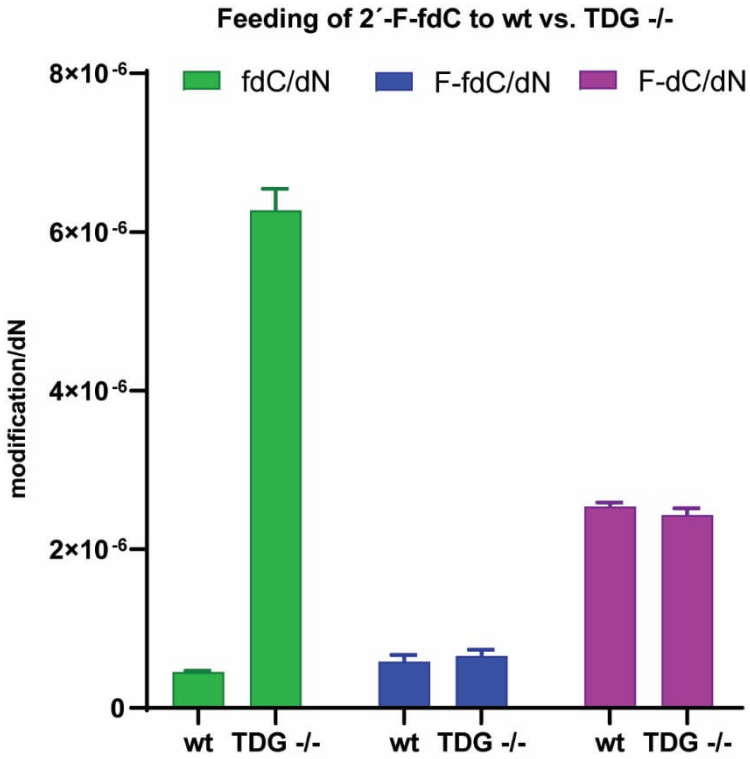


Figure SI-6: Levels of natural fdC (left, green), incorporated modified 2´-F-fdC (**1**) in the genome (middle, blue) and of the deformylation product 2´-F-dC (**2**, right, magenta) of a wildtype and Tdg -/- cell line. Levels of genomic natural fdC increase dramatically in absence of Tdg activity, while the levels of fluorinated-derivatives (**1** and **2**) stay constant. The error bars represent the standard deviation.

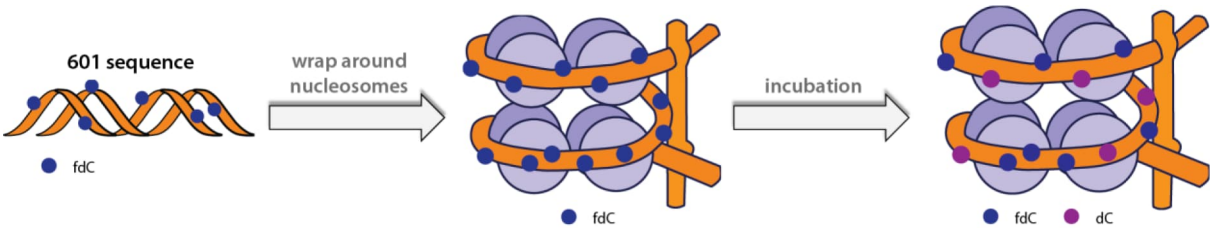


Figure SI-7A: Experimental workflow of PCR-produced 601 Widom sequence[1] (shown below) and its incubation with nucleosomes, to determine the involvement of nucleosomes on deformylation of fdC.

601 sequence:

5´-ATCGATGTATATATCTGACACGTGCCTGGAGACTAGGGAGTAATCCCCTTGGCGGTTAAAA CGCGGGGGACAG-3´

Forward Primer: 5´-ATCGATGTATATATCTGACACGTGC-3´

Reverse Primer: 5´-ATCAGAATCCCGGTGCCGAG-3´

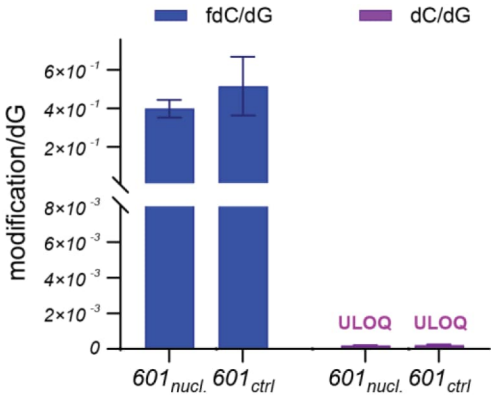


Figure SI-7B: Levels of fdC and dC found in the 601 sequence with or without incubation with nucleosomes. Error bars represent standard deviation of three technical replicates.

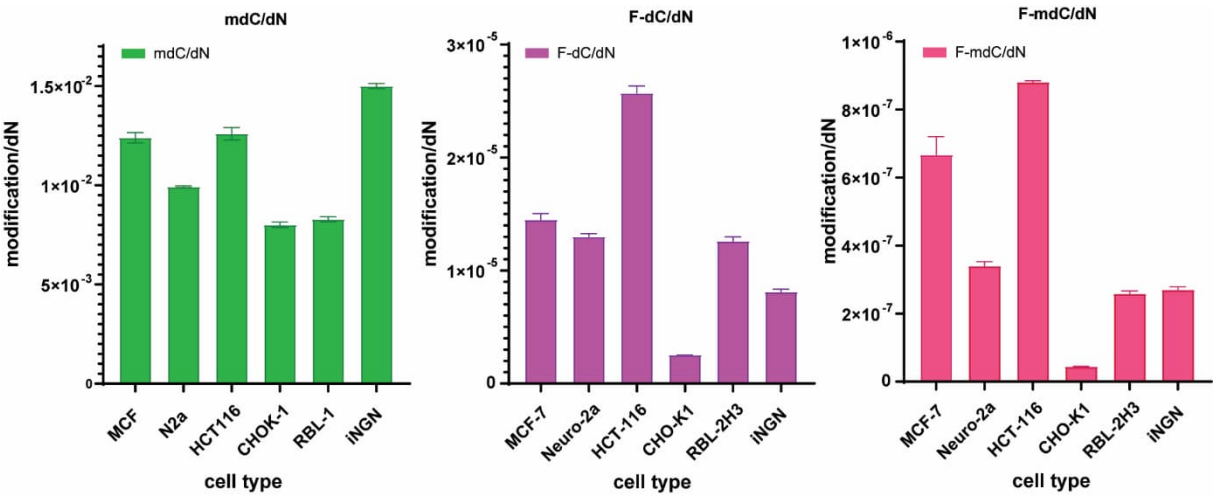


Figure SI-8A: Levels of incorporated natural mdC (green, left), 2′-F-dC (**2**, purple, middle) and 2′-F-mdC (**9**, pink, right) in different cell lines. Error bars represent standard deviation of three biological replicates.

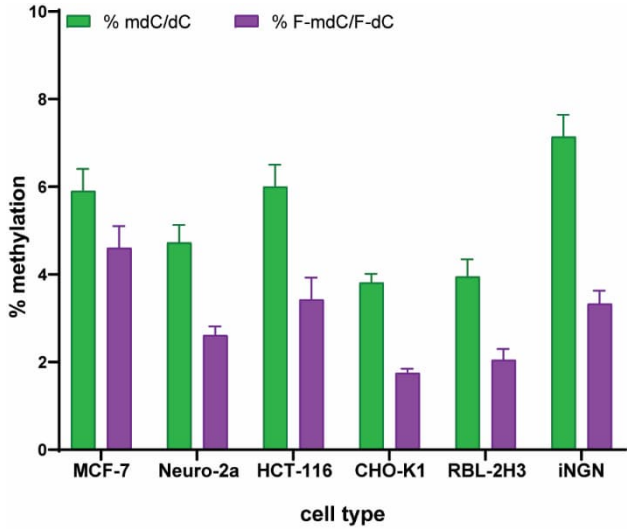


Figure SI-8B: Methylation levels of natural dC and F-dC in different cell lines. Error bars represent standard deviation of three biological replicates.

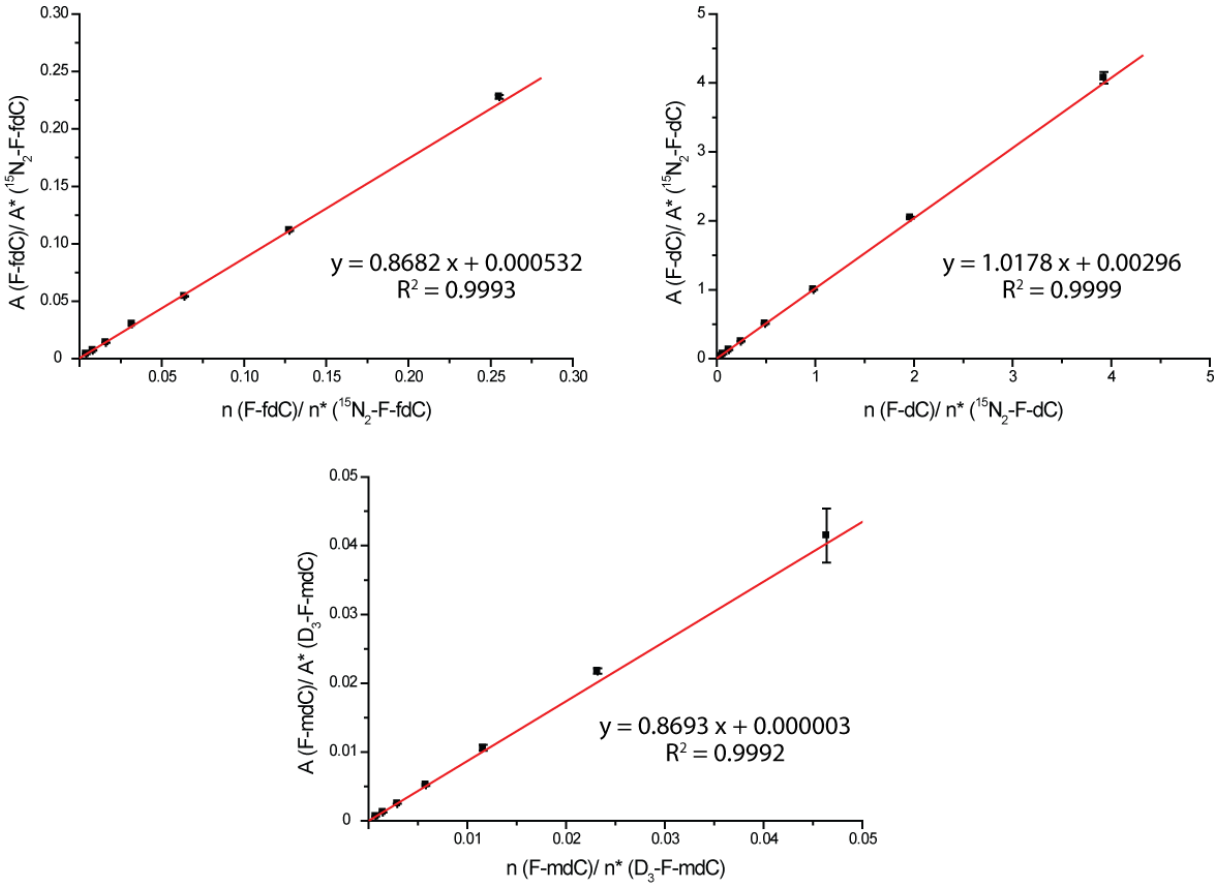


Figure SI-9: Internal calibration curves for the exact quantification of 2'-F-fdC (**1**), 2'-F-dC (**2**) and 2'-F-mdC (**9**) with the corresponding linear equation and coefficient of determination.[2]

Methods and Materials

Chemical Synthesis

Unless noted otherwise, all chemical transformations were performed using flame- or oven dried glassware under an atmosphere of argon. Chemicals were purchased from *Sigma-Aldrich*, *TCI*, *Fluka*, *ABCR*, *Carbosynth* or *Acros Organics* and used without further purification. Technical grade solvents were distilled prior to extraction or chromatography. Reactions and column chromatography fractions were monitored by thin-layer chromatography (TLC) on silica gel F_{254}TLC plates from Merck KGaA. Visualization of the developed TLC plates was achieved through UV-absorption or through staining with Hanessian's stain or $KMnO_4$ stain. Flash column chromatography was performed on Geduran®Si60 (40-63 µm) silica gel from Merck KGaA applying slight nitrogen pressure. NMR spectra were recorded on *Bruker Avance III HD 400 (*400 MHz), *Varian NMR-System 600* (600 MHz) and *Bruker Avance III HD* with *Cryo-Kopf 800* (800 MHz) spectrometers. ^{1}H-NMR shifts were calibrated to the residual solvent resonances: $CDCl_3$ (7.26 ppm), DMSO-d_6 (2.50ppm) and D_2O (4.79 ppm). ^{13}C-NMR shifts were calibrated to the residual solvent: $CDCl_3$ (77.16 ppm), DMSO-d_6 (39.52 ppm). All NMR spectra were analysed using the program MestReNova 10.0.1 from Mestrelab NMR data are reported as follows: chemical shift (multiplicity, coupling constants where applicable, number of hydrogens, assignment). Splitting is reported with the following symbols: s = singlet, d = doublet, t = triplet, dd = doublet of doublets, dt = doublet of triplets, ddd = doublet of doublets of doublets, m = multiplet. Whenever a signal is not clearly distinguishable from related nuclides, roman numerals (e.g. "I", "IV") are attached to indicate interchangeability. Research S. L. Low resolution mass spectra were measured on a LTQ FT-ICR by Thermo Finnigan GmbH. High resolution mass spectra were measured by the analytical section of the Department of Chemistry of Ludwig-Maximilians-Universität München on aMAT 90 (ESI) from Thermo Finnigan GmbH. IR spectra were recorded on a PerkinElmer Spectrum BX II FT-IR system. Substances were applied as a film or directly as solids on the ATR unit. Analytical RP-HPLC was performed on an analytical HPLC Waters Alliance(2695 Separation Module, 2996 Photodiode Array Detector) equipped with the column Nucleosil120-2 C18 from Macherey Nagel applying an eluent flow of 0.5 mL/min. Preparative RP-HPLC was performed on HPLC Waters Breeze (2487 Dual λ Array Detector, 1525 Binary HPLC Pump) equipped with the column VP 250/32 C18 from Macherey Nagel. A flowrate of 5mL/min was applied. HPLC grade MeCN was purchased from VWR.

Cell culture

DMEM high glucose (Sigma Aldrich) containing 10% heat-inactivated FBS (Gibco), 100 U/mL Penicillin-Streptomycin (Gibco), 2 mM L-glutamine were used to culture Neuro-2a, RBL-2H3 ad MCF-7 cell lines.

DMEM high glucose and Ham's Nutrient Mixture F12 (Sigma Aldrich) mixture 50/50 containing 100 U/mL Penicillin-Streptomycin (Gibco), 2 mM L-glutamine were used to culture CHO-K1 cells.

Minimum Essential Medium Eagle (MEM) (Sigma Aldrich) containing 10% heat-inactivated FBS (Gibco), 100 U/mL Penicillin-Streptomycin (Gibco), 2 mM L-glutamine and were used to culture HCT-116 cells.

iNGN cell line[3] at somatic cell state was cultured in "E7 media" consisted of DMEM high glucose (Sigma Aldrich) and Ham's Nutrient Mixture F12 (Sigma Aldrich) 50/50 mixture supplemented with 100 U/mL Penicillin-Streptomycin (Gibco), 64mg/L L -Ascorbic acid 2-phosphate, 0.1mM Nicotinamide (Sigma Aldrich), 78nM Sodium selenite (Sigma Aldrich), 0.64g/L NaCl (Sigma Aldrich), 20 µg/mL Insulin in 25mM HEPES buffer pH 8.2 (Sigma Aldrich), 10 µg/mL Holo-transferrin (Sigma Aldrich). Additionally, with freshly added growth factors 20 ng/mL FGF-2 and TGF-β 10 ng/mL (MACS) making "E9" media. After every splitting the cells were treated with the adherence factor – Thiazovivin (TOCRIS) at 1uM added freshly to the plate. During differentiation the cells were cultured in E7 medium complemented by doxycycline 0,5 µg/mL (Sigma Aldrich) for the first 72 hours. After 72nd hour of differentiation the medium was changed into Neurobasal A (Gibco) + 2% NeuroBrew21 (MACS). The iNGNs were always plated on Geltrex coating (Thermo Fisher) according to manufactures protocol.

mES cells were cultured in three different media – slitting, maintenance (a2i) and priming (C/R) media. The maintenance medium consisted of DMEM high glucose (Sigma Aldrich), 10% heat-inactivated FBS (Gibco), 0.1mM β-mercaptoethanol (Sigma Aldrich), 2mM L-Alanyl-L-Glutamine 200mM (Gibco), 1x Nonessential amino acids 100x (Sigma Aldrich) and 100 U/mL Penicillin-Streptomycin (Gibco). The culture medium consists of the splitting medium supplemented with LIF 1000U/mL (ORF Genetics), 3.0 µM CGP 77675 (Sigma Aldrich) and 3.0 µM GSK3 inhibitor CHIR-99021 (Sigma Aldrich). The priming media consisted of the splitting media supplemented with LIF 1000U/mL (ORF Genetics), 2.5 µM tankyrase inhibitor IWR-1 endo (Sigma Aldrich) and 3.0 µM CHIR-99021 (Sigma Aldrich). The cells were maintained in a2i medium and primed for 3.0 days via plating in C/R medium.

Metabolic-labelling experiments were conducted by plating cells in their dedicated medium containing 350μM of F-fdC for 3 days. mESCs were fed during priming without medium change until the moment of harvesting. The pulse-chase experiment using iNGN cells was performed so that the cells were fed dT+12 and F-fdC on the day of induction and the medium with nucleosides was replaced after 24hours and from the 48th hour the chase has started. The harvesting timepoints from the first hour of induction were: 48h, 52h, 56h, 60h, 68h, 72h and 76h.

The cells were counted before plating and prior to harvesting (at timepoints 0 h and 72 h) using Countess II automatic cell counter (LifeTechnologies. ThermoFisher) via trypan blue staining. The values were recorded and the doubling rates calculated according to the following equation:

$$doubling\ time = \frac{72 * \log(2)}{log_2(F) - log_2(I)}$$

F = final number of cells, I = initial number of cells.

Cell Harvesting and Isolation of genomic DNA

All samples were washed with PBS (Sigma Aldrich) once before harvesting and lysed directly in the plates by adding RLT buffer (Qiagen) supplemented with 400μM 2,6-di-tert-butyl-4-methylphenol (BHT) and desferoxamine mesylate (DM). Next, DNA was sheered in MM400 bead mill (Retsch) at 30 Hz for 1 min in 2mL microcentrifuge tubes (Eppendorf) with 5mm diameter stainless steel beads (one per tube) and centrifuged at 21000rcf for 5 minutes. Genomic DNA (gDNA) was extracted using Zymo Quick gDNA mini-prep® kit according to the protocol with an addition of RNAse A treatment (Qiagen) at 0.2mg/mL in Genomic Lysis Buffer – 400μL of the solution transferred directly on the column and incubated for 15 minutes. All samples were eluted in DNAse-free ddH_2O with 20μM BHT, the concentration of gDNA was measured on Nanodrop.

DNA digestion

While investigating the deformylation levels using F-fdC as a metabolic label, the following conditions were used for the gDNA digestion procedure: 1 – 10 μg of genomic DNA in 35 μL H_2O were digested as follows: 2 μL of Degradase plus (*Zymo Research*), 5 μL of Degradase plus reaction buffer (*Zymo Research*), 0.5 μL of Benzonase nuclease (*Merck, Serratia marcescens*) and a specific amount of isotopically labelled internal standards were added. The

mixture was incubated for 4 h at 37 °C and then stored at -20 °C. Prior to LC-MS/MS analysis, samples were filtered by using an AcroPrep Advance 9 filter plate 0.2 µm Supor (*Pall Life Science*).

LC/MS-MS analysis of DNA samples[4]

Quantitative UHPLC-MS/MS analysis of digested DNA samples was performed using an Agilent 1290 UHPLC system equipped with a UV detector and an Agilent 6490 triple quadrupole mass spectrometer. Nucleosides of interest were quantified using the stable isotope dilution technique.[2] The source-dependent parameters were as follows: gas temperature 80 °C, gas flow 15 L/min (N_2), nebulizer 30 psi, sheath gas heater 275 °C, sheath gas flow 11 L/min (N2), capillary voltage 2.500 V in the positive ion mode, capillary voltage −2.250 V in the negative ion mode and nozzle voltage 500 V. The fragmentor voltage was 380 V/ 250 V. Delta EMV was set to 500 (positive mode) and 800 (negative mode). Compound-dependent parameters are summarized in Supplementary Tables 1-3. Chromatography was performed by a Poroshell 120 SB-C8 column (Agilent, 2.7 µm, 2.1 mm × 150 mm) at 35 °C using a gradient of water and MeCN, each containing 0.0085% (v/v) formic acid, at a flow rate of 0.35 mL/min: 0 →4 min; 0 →3.5% (v/v) MeCN; 4 →7.9 min; 3.5 →5% MeCN; 7.9 →8.2 min; 5 →80% MeCN; 8.2 →11.5 min; 80% MeCN; 11.5 →12 min; 80 →0% MeCN; 12 →14 min; 0% MeCN. The autosampler was cooled to 4 °C. The injection volume was amounted to 39 µL.

Quantification of nucleosides in DNA samples

Nucleosides were quantified using internal calibration curves and the stable isotope dilution technique as described in the literature.[2] Used internal calibration curves for F-nucleosides are shown in Supplementary Figure 4.

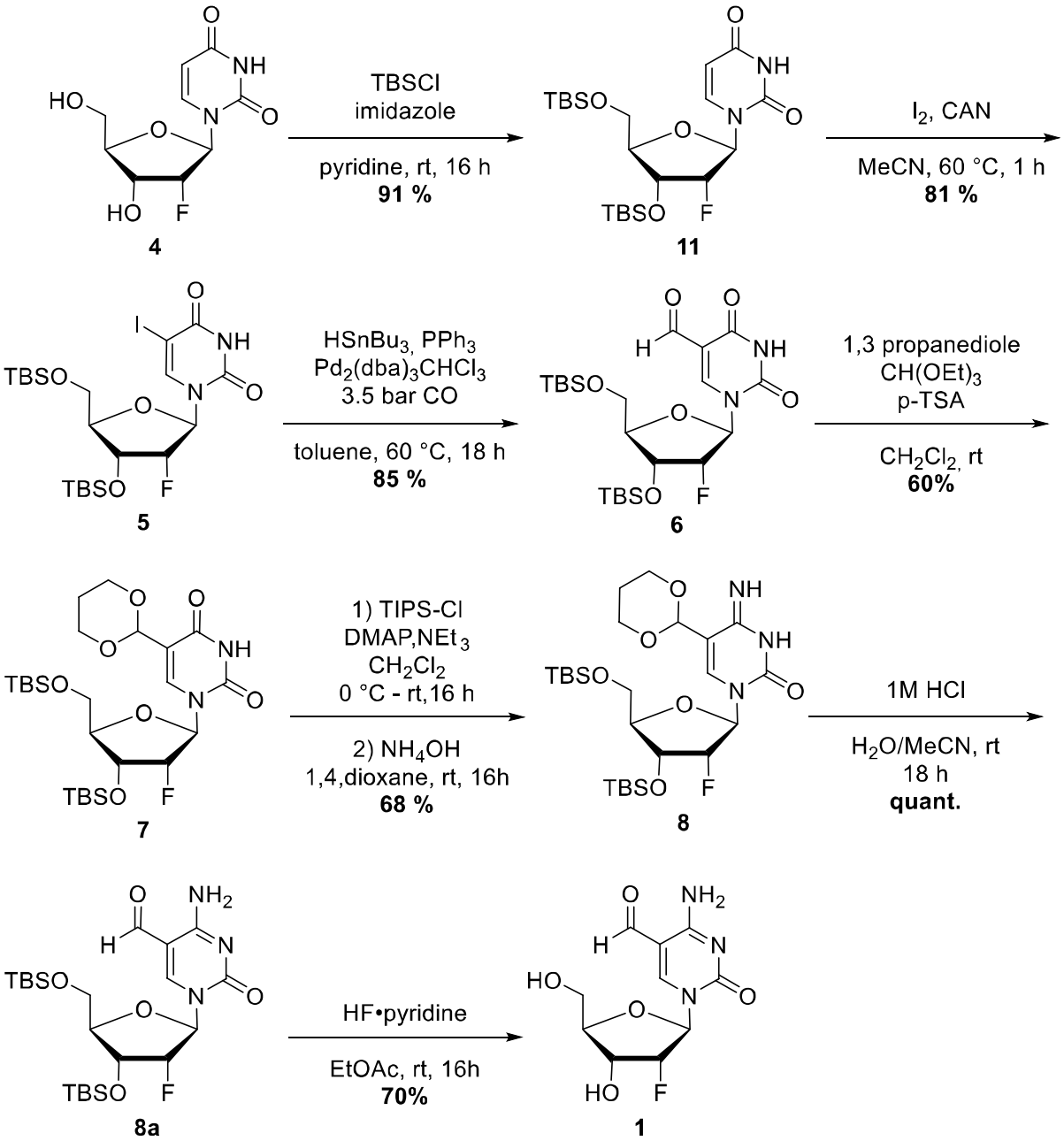
Soluble nucleoside pool extraction and purification

The soluble pool extraction protocol started with a cell harvest using typical cell passaging trypsinisation technique rather than a usual cell lysis method (using TrypLE reagent, Gibco). The cells were counted with trypan blue via Countess II FL Automated Cell Counter, Life Technologies (ThermoFisher Scienitfic) and pelleted via centrifugation at 260g for 3min. The pellet was resuspended with the ice-cold mixture of acetonitrile and water (1:1 ratio), 1mL per $1x10^6$cells and gently vortexed. The cell suspension was incubated on ice for 15min and centrifuged at 16 000g for 10min at 4ºC. The supernatant containing the cytosolic fraction was

collected and lyophilised, while the cell pellet was treated with the lysis buffer for subsequent gDNA isolation.

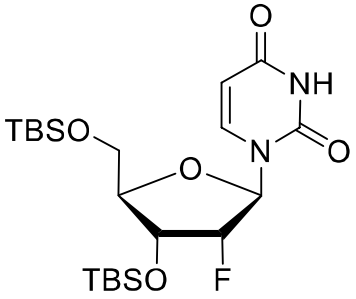
Samples were dissolved in H_2O (pH = 4, 100 µL H_2O for soluble pool from 400 000 cells). Supel-Select HBL 60 mg columns were equilibrated with 1.5 mL MeOH and 3 mL H_2O (pH = 4). Samples were loaded the surface of the column, washed with 4 mL H_2O (pH = 4). Cartridges were dried under high vacuum for 15 min. Sample was eluted with 1.5 mL MeOH/MeCN = 1:1. Eluent was lyophilized, dissolved in H_2O and analysed by UHPLC-MS/MS as described for genomic DNA samples.[2]

Synthetic procedures



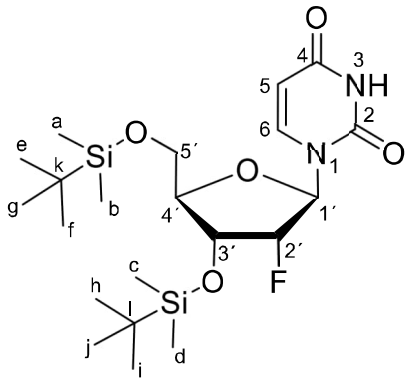
Scheme SI-1: Synthesis pathway towards 2'-F-fdC (**1**).

***Synthesis* of O3´,5´-Bis(tertbutyldimethylsilyl)-2´-fluorouridine (11)**



11

According to a modified literature [5] procedure 2'-deoxy-2'-fluoruridine (2.0 g, 8.1 mmol, 1.0 equiv.), TBS-Cl (3.7 g, 24.4 mmol, 3.0 equiv.) and imidazole (2.5 g, 36.5 mmol, 4.5 equiv.) were combined in a Schlenk flask an dissolved in pyridine (32 mL). The resulting reaction mixture was stirred at room temperature for 18 h. The reaction was stopped by addition of $NaHCO_3$ (sat. aq., 150 mL), extracted with dichloromethane (3 x 40 mL). The combined organic layers were washed with brine (150 mL), dried over Na_2SO_4 and the solvent was removed *in vacuo*. Residual pyridine was co-evaporated with toluene yielding a colourless crude. Purification by flash column chromatography (SiO_2 hexanes/ethyl acetate 2:1) yielded target compound **11** as a colourless solid (3.8 g, 8.0 mmol, 99 %).



R_f = 0.48 (CH_2Cl_2/MeOH 9:1)

^{1}H-NMR (599 MHz, $CDCl_3$, ppm): δ = 8.06 (s, 1H, N3-H), 7.91 (s, 1H, $^3J_{H6\text{-}H5}$ = 8.0 Hz, $^4J_{H6\text{-}H1´}$ = 0.9 Hz, C6-H), 6.06 (dd, $^3J_{H1´\text{-}F2´}$ = 15.2 Hz, $^3J_{H1´\text{-}H2´}$ = 2.0 Hz, 1H, C1'-H), 5.68 (dd, $^3J_{H5\text{-}H6}$ = 8.0 Hz, $^4J_{H5\text{-}H3}$ = 0.9 Hz, C5-H), 4.77 (ddd, $^2J_{H2´\text{-}F2´}$ = 52.4 Hz, $^3J_{H2´\text{-}H1´}$ = 2.0 Hz, $^3J_{H2´\text{-}H3´}$ = 4.3 Hz, 1H, C2'-H), 4.29 (ddd, $^3J_{H3´\text{-}F2´}$ = 18.7 Hz, $^3J_{H3´\text{-}H2´}$ = 4.2 Hz, $^3J_{H3´\text{-}H4´}$ = 7.2 Hz, 1H, C3'-H), 4.08

(dd, $^{3}J_{H4'-H3'}$ = 7.2 Hz, $^{3}J_{H4'-H5'}$ = 1.6 Hz, 1H, C4'-H), 4.05 (dd, $^{2}J_{H5'a-H5'b}$ = 11.7 Hz, $^{3}J_{H5'a-H4'}$ = 2.0 Hz, 1H, C5'-H_a)^{I}, 3.78 (dd, $^{2}J_{H5'b-H5'a}$ = 11.8 Hz, $^{3}J_{H5'b-H4'}$ = 1.7 Hz, 1H, C5'-H_b)^{I} , 0.93 (s, 9H, C_{e-g}-H3)^{II}, 0.91 (s, 9H, C_{h-j}-H3)^{II}, 0.12 (s, 6H, $C_{a,b}$-H3)^{III}, 0.12 (s, 3H, C_c-H3)^{III}, 0.11 (s, 3H, C_d-H3)^{III}.

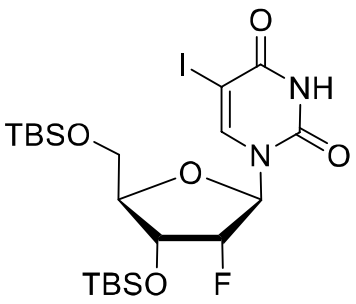
^{13}C-NMR (101 MHz, $CDCl_3$, ppm): δ = 162.6 (C4), 149.9 (C2), 139.9 (C6), 102.5 (C5), 93.2 (d, $^{1}J_{C2'-F2'}$ = 193.5 Hz, C2'), 87.9 (d, $^{2}J_{C1'-F2'}$ = 33.8 Hz, C1'), 84.0 (C4'), 68.8 (d, $^{2}J_{C3'-F2'}$ = 16.3 Hz, C3'), 60.9 (C5'), 26.1 (3C, C_{e-g})^{I} , 25.8 (3C, C_{h-j})^{I}, 18.6 (C_k)^{II}, 18.3 (C_l)^{II}, -4.5 (C_a)^{III}, -4.9 (C_b)^{III}, -5.3 (C_c)^{III}, -5.3 (C_d)^{III}.

^{19}F-NMR (377 MHz, $CDCl_3$, ppm): δ = -202.6 (ddd, $^{2}J_{F2'-H2'}$ = 52.5 Hz, $^{3}J_{F2'-H3'}$ = 18.7 Hz, $^{3}J_{F2'-H1'}$ = 15.2 Hz, C2'-F).

HRMS (ESI+): calc. for $C_{21}H_{40}FN_2O_5Si_2^+$ $[M+H]^+$: 475.2454, found: 475.2453.
HRMS (ESI-): calc. for $C_{21}H_{38}FN_2O_5Si_2^-$ $[M-H]^-$: 473.2309, found: 473.2319.

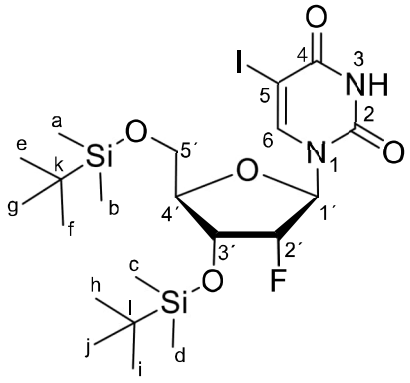
IR (ATR): ν /cm^{-1} = 3390 (w), 3261 (w), 2593 (w), 2928 (w), 2857 (w), 1759 (w), 1712 (s), 1707 (s), 1695 (vs), 1628 (w), 1455 (m), 1251 (m), 1158 (m), 1120 (s), 994 (m), 978 (m), 880 (ms), 863 (m), 828 (vs), 803 (s) 760 (s), 665 (m).

***Synthesis of* O3´,5´-Bis(tertbutyldimethylsilyl)-2´-fluor-5-iodouridine (5)**



5

TBS-protected uridine **11** (2.1 g, 4.4 mmol, 1.0 equiv.) was dissolved in dry acetonitrile (200 mL), ceric ammonium nitrate (5.3 g, 9.6 mmol, 2.2 equiv.) and iodine (2.4 g, 9.6 mmol, 2.2 equiv.) were added, and the resulting reaction mixture was stirred for one hour at 60 °C. Upon complete conversion of starting material, the reaction was stopped by addition of $NaHCO_3$ (sat. aq., 100 mL) and $Na_2S_2O_3$ (sat. aq., 100 mL). Reaction mixture was filtered over a plug of SiO_2. The obtained filtrate was concentrated *in vacuo*, the aqueous residue was extracted with dichloromethane (4 x 40 mL). The combined organic layers were washed with brine (150 mL), dried over Na_2SO_4 volatiles were removed under reduced pressure. Resulting crude product was purified by flash column chromatography (SiO_2, hexanes/ethyl acetate 4:1) yielding **5** as a colourless foam (2.1 g, 3.5 mmol, 81 %).



R_f = 0.75 (hexanes/ethyl acetate 2:1)

^{1}H-NMR (599 MHz, $CDCl_3$, ppm): δ = 8.16 (s, 1H, N3-H), 7.98 (s, 1H, C6-H), 6.09 (dd, $^3J_{H1´-F2´}$ = 14.0 Hz, $^3J_{H1´-H2´}$ = 4.1 Hz, 1H, C1'-H), 4.82 (dt, $^2J_{H2´-F2´}$ = 52.6 Hz, $^3J_{H2´-H1´,H3´}$ = 4.4 Hz, 1H, C2'-H), 4.29 (dt, $^3J_{H3´-F2´}$ = 11.3 Hz, $^3J_{H3´-H2´,H4´}$ = 4.8 Hz, 1H, C3'-H), 4.08 (dd, $^3J_{H4´-H3´}$ = 4.9 Hz, $^3J_{H4´-H5´}$ = 2.1 Hz, 1H, C4'-H), 3.98 (dd, $^2J_{H5´a-H5´b}$ = 11.8 Hz, $^3J_{H5´a-H4´}$ = 1.9 Hz, 1H, C5'-H_a)^I, 3.77

(dd, $^{2}J_{H5'b\text{-}H5'a}$ = 11.8 Hz, $^{3}J_{H5'b\text{-}H4'}$ = 2.2 Hz, 1H, C5'-H_b)^{I} , 0.97 (s, 9H, $C_{e\text{-}g}$-H_3)^{II}, 0.91 (s, 9H, $C_{h\text{-}j}$-H_3)^{II}, 0.18 (s, 3H, C_a-H_3)^{III}, 0.17 (s, 3H, C_b-H_3)^{III} 0.13 (s, 3H, C_c-H_3)^{III}, 0.11 (s, 3H, C_d-H_3)^{III}.

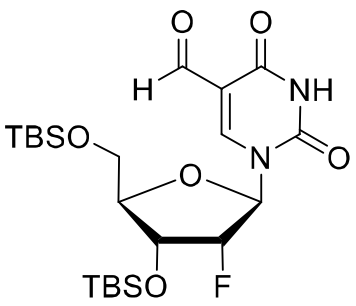
^{13}C-NMR (101 MHz, $CDCl_3$, ppm): δ = 159.5 (C4), 149.6 (C2), 144.1 (C6), 92.5 (d, $^{1}J_{C2'\text{-}F2'}$ = 195.7 Hz, C2'), 87.5 (d, $^{2}J_{C1'\text{-}F2'}$ = 33.1 Hz, C1'), 85.5 (C5), 69.9 (d, $^{2}J_{C3'\text{-}F2'}$ = 15.4 Hz, C3'), 69.2 (C4´), 62.0 (C5'), 26.5 (3C, $C_{e\text{-}g}$)^{I} , 25.8 (3C, $C_{h\text{-}j}$)^{I}, 18.8 (C_k)^{II}, 18.3 (C_l)^{II}, -4.5 (C_a)^{III}, -4.9 (2C, $C_{b,c}$)^{III}, -5.3 (C_d)^{III}.

^{19}F-NMR (377 MHz, $CDCl_3$, ppm): δ = -205.7 (dt, $^{2}J_{F2'\text{-}H2'}$ = 52.6 Hz, $^{3}J_{F2'\text{-}H1',H3'}$ = 12.6 Hz, C2'-F).

HRMS (ESI+): calc. for $C_{21}H_{39}FIN_2O_5Si_2^+$ $[M+H]^+$: 601.1421, found: 601.1419.
HRMS (ESI-): calc. for $C_{21}H_{37}FIN_2O_5Si_2^-$ $[M-H]^-$: 599.1275, found: 599.1292.

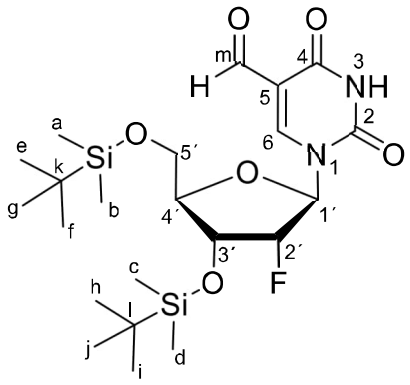
IR (ATR): ν /cm^{-1} = 3171(w), 3067 (w), 2953 (w), 2928 (m), 2856 (w), 1690 (vs), 1606 (m), 1471 (w), 1426 (w), 1407 (w), 1360 (w), 1253 (s), 1166 (m), 1127 (s), 1072 (s), 996 (w), 680 (w), 889 (m), 858 (m), 802 (vs), 778 (vs), 670 (m).

***Synthesis of* O3´,5´-Bis(tertbutyldimethylsilyl)-2´-fluor-5-formyluridine (6)**



6

According to a modified literature procedure[5] Pd_2dba_3•$CHCl_3$ (0.14 g, 0.13 mmol, 0.1 equiv.), PPh_3 (0.21 g, 0.80 mmol, 0.6 equiv.) and **5** (0.80 g, 1.3 mmol, 1.0 equiv.) were charged to an autoclave vessel and dissolved in dry toluene (26 mL). The autoclave apparatus was sealed, the reaction mixture was purged with carbon monoxide (3 x 3.5 bar) and subsequently heated to 60 °C. A solution of $HSnBu_3$ (0.47 g, 0.43 mL, 1.6 mmol, 1.2 equiv.) in toluene (0.4 mL) was added via a syringe pump (55 µL/h) over a period of 15 h. After complete addition, reaction mixture was let to cool to room temperature and filtered over a plug of silica (eluted with 350 mL ethyl acetate). The resulting crude product was purified with flash column chromatography (SiO_2, 10 w% K_2CO_3, hexanes/ethyl acetate 4:1) yielding the formylated compound **6** as an colourless foam (560 mg, 1.12 mmol, 84 %).



R_f = 0.75 (hexanes/ethyl acetate 2:1)

^{1}H-NMR (599 MHz, $CDCl_3$, ppm): δ = 10.22 (s, 1H, C_m-H), 8.46 (s, 1H, N3-H), 8.42 (s, 1H, C6-H), 6.12 (dd, $^3J_{H1´-F2´}$ = 13.9 Hz, $^3J_{H1´-H2´}$ = 4.0 Hz, 1H, C1'-H), 4.89 (dt, $^2J_{H2´-F2´}$ = 52.5 Hz, $^3J_{H2´-H1´,H3´}$ = 4.3 Hz, 1H, C2'-H), 4.31 (dt, $^3J_{H3´-F2´}$ = 11.2 Hz, $^3J_{H3´-H2´,H4´}$ = 4.8 Hz, 1H, C3'-H), 4.16-4.12 (m, 1H, C4'-H), 3.99 (dd, $^2J_{H5´a-H5´b}$ = 11.9 Hz, $^3J_{H5´a-H4´}$ = 2.1 Hz, 1H, C5'-H_a)^l, 3.78 (dd,

$^{2}J_{H5'b\text{-}H5'a}$ = 11.8 Hz, $^{3}J_{H5'b\text{-}H4'}$ = 2.4 Hz, 1H, C5'-H_b)^{I} , 0.93 (s, 9H, $C_{e\text{-}g}$-H_3)^{II}, 0.91 (s, 9H, $C_{h\text{-}j}$ -H_3)^{II}, 0.15 (s, 6H, $C_{a,b}$-H_3)^{III}, 0.13 (s, 3H, C_c-H_3)^{III}, 0.11 (s, 3H, C_d-H_3)^{III}.

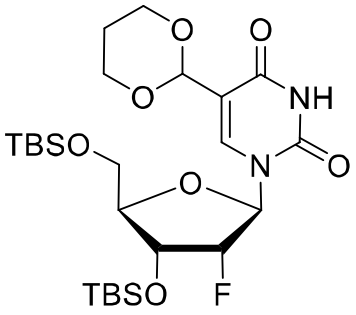
^{13}C-NMR (101 MHz, $CDCl_3$, ppm): δ = 185.5 (C_m), 161.1 (C4), 149.0 (C2), 145.3 (C6), 111.8 (C5), 92.7 (d, $^{1}J_{C2'\text{-}F2'}$ = 196.4 Hz, C2'), 88.3 (d, $^{2}J_{C1'\text{-}F2'}$ = 33.4 Hz, C1'), 86.0 (C4´), 70.0 (d, $^{2}J_{C3'\text{-}F2'}$ = 15.4 Hz, C3'), 61.9 (C5´), 26.2 (3C, $C_{e\text{-}g}$)^{I} , 25.8 (3C, $C_{h\text{-}j}$)^{I}, 18.7 (C_k)^{II}, 18.3 (C_l)^{II}, -4.5 (C_a)^{III}, -4.9 (C_b)^{III}, -5.3 (C_c)^{III} -5.4 (C_d)^{III}.

^{19}F-NMR (377 MHz, $CDCl_3$, ppm): δ = -205.3 (dt, $^{2}J_{F2'\text{-}H2'}$ = 52.2 Hz, $^{3}J_{F2'\text{-}H1',H3'}$ = 12.2 Hz, C2'-F).

HRMS (ESI+): calc. for $C_{22}H_{40}FN_2O_6Si_2^+$ $[M+H]^+$: 503.2403, found: 503.2400.

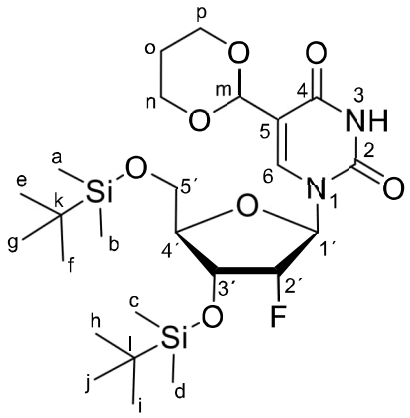
IR (ATR): ν /cm^{-1} = 2949 (w), 2926 (w), 2907 (w), 2889 (w), 2169 (m), 1692 (s), 1599 (m), 1470 (m), 1379 (w), 1361 (w), 1254 (m), 1233 (w), 1121 (m), 1073 (m), 997 (w), 982 (w), 830 (vs), 777 (vs), 673 (m), 658 (w).

***Synthesis of* O3´,5´-Bis(tertbutyldimethylsilyl)-2´-fluor-5-(1,3-dioxan-2-yl)uridine (7)**



7

Nucleoside **6** (310 mg, 0.617 mmol, 1.0 equiv.) was dissolved in dry dichloromethane (12 mL) and cooled to 0 °C. At 0 °C 1,3-propanediol (188 mg, 0.180 mL, 2.47 mmol, 4.0 equiv.), triethylorthoformiate (183 mg, 0.210 mL, 1.23 mmol, 2.0 equiv.) and p-TSA (24.0 mg, 0.123 mmol, 0.2 equiv.) were added. The reaction mixture was stirred for 30 min at 0 °C and 60 min at room temperature. Reaction was quenched by addition of H_2O (25 mL) and extracted with CH_2Cl_2 (3 x 30 mL). Combined organic layers were washed with brine (2 x 80 mL) and dried over Na_2SO_4. After all volatiles were removed under reduced pressure, crude was purified via flash column chromatography (SiO_2, hexanes/ethyl acetate 2:1) to yield **7** as a colourless foam (218 mg, 0.389 mmol, 63 %).



R_f = 0.25 (hexanes/ethyl acetate 2:1)

^{1}H-NMR (599 MHz, $CDCl_3$, ppm): δ = 8.06 (s, 1H, N3-H), 7.66 (d, $^4J_{H6\text{-}Hm}$ = 0.8 Hz, 1H, C6-H), 5.97 (dd, $^3J_{H1´\text{-}F2´}$ = 15.7 Hz, $^3J_{H1´\text{-}H2´}$= 4.3 Hz, 1H, C1'-H), 5.52 (s, 1H, CH_m), 4.91 (dt, $^2J_{H2´\text{-}F2´}$ = 53.3 Hz, $^3J_{H2´\text{-}H1´;H3´}$ = 4.6 Hz, 1H, C2'-H), 4.35 (dt, $^3J_{H3´\text{-}F2´}$ = 10.7 Hz, $^3J_{H3´\text{-}H2´;H4´}$ = 5.2 Hz, 1H,

C3'-H), 4.20-4.15 (m, 2H, C_n-H_2)^{I} , 4.01 (dp, $^{3}J_{H4'-H3'}$ = 5.2 Hz, $^{3}J_{H4'-H5'a;H5'b}$ = 1.7 Hz, 1H, C4'-H), 3.98-3.91 (m, 2H, C_p-H_2)^{I}, 3.90 (dd, $^{2}J_{H5'a-H5'b}$ = 11.7 Hz, $^{3}J_{H5'a-H4'}$ = 2.3 Hz, 1H, C5'-H_a)^{II}, 3.76 (dd, $^{3}J_{H5'b-H5'a}$ = 11.7 Hz, $^{3}J_{H5'b-H4'}$ = 3.4 Hz, 1H, $C_{5'}$-H_b)^{II}, 2.16-2.08 (m, 1H, C_o-H_2)^{III}, 1.43-1.38 (m, 1H, C_o-H_2)^{III}, 0.93 (s, 9H, C_{e-g}-H_3)^{IV}, 0.91 (s, 9H, C_{h-j} -H_3)^{IV}, 0.12 (s, 3H, C_a-H_3)^{V}, 0.11 (s, 6H, $C_{b;c}$-H_3)^{V}, 0.11 (s, 3H, C_d-H_3)^{V}.

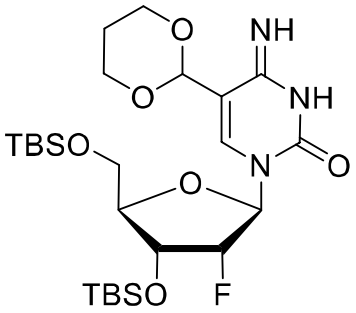
^{13}C-NMR (101 MHz, $CDCl_3$, ppm): δ = 161.1 (C4), 149.6 (C2), 169.7 (C6), 113.2 (C5), 95.5 (C_{acetal}) 91.8 (d, $^{1}J_{C2'-F2'}$ = 194.5 Hz, C2'), 88.7 (d, $^{2}J_{C1'-F2'}$ = 33.4 Hz, C1'), 86.0 (C4´), 69.9 (d, $^{2}J_{C3'-F2'}$ = 15.3 Hz, C3'), 68.7 (C_n)^{I}, 67.7 (C_p)^{I}, 62.3 (C5´), 26.2 (3C, C_{e-g})^{II} , 25.9 (3C, C_{h-j})^{II}, 18.7 (C_k)^{III}, 18.4 (C_l)^{III}, -4.6 (C_a) ^{VI}, -4.9 (C_b)^{IV}, -5.1 (C_c) ^{VI} -5.4 (C_d)^{IV}.

^{19}F-NMR (377 MHz, $CDCl_3$, ppm): δ = -204.7 (ddd, $^{2}J_{F2'-H2'}$ = 53.4 Hz, $^{3}J_{F2'-H3'}$ = 15.6 Hz, $^{3}J_{F2'-H1'}$ = 11.1 Hz C2'-F).

HRMS (ESI+): calc. for $C_{25}H_{46}FN_2O_7Si_2^+$ $[M+H]^+$: 561.2822, found: 561.2824.
HRMS (ESI-): calc. for $C_{25}H_{44}FN_2O_7Si_2^-$ $[M-H]^-$: 559.2677, found: 559.2690.

IR (ATR): ν /cm^{-1} = 3191 (w), 2954 (w), 2928 (w), 2887 (w), 2856 (w), 1697 (s), 1470 (m), 1463 (m), 1362 (m), 1257 (m), 1127 (m), 1092 (s), 1004 (m), 966 (m), 860 (m), 802 (vs), 777 (vs), 724 (m), 671 (m).

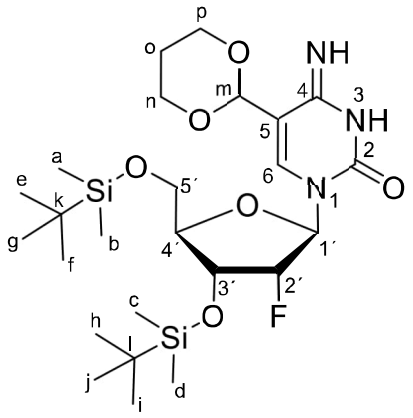
***Synthesis of* O3´,5´-Bis(tertbutyldimethylsilyl)-2´-fluor-5-(1,3-dioxan-2-yl)cytidine (8)**



8

Nucleoside **7** (370 mg, 0.66 mmol, 1.0 equiv.) and DMAP (16.0 mg, 0.13 mmol, 0.2 equiv.) were dissolved in dry dichloromethane (7 mL), triethylamine (0.9 mL, 0.67 g, 6.6 mmol, 10 equiv.) was added. Reaction mixture was cooled to 0 °C. At 0 °C triisopropylsulfonylchloride (300 mg, 0.99 mmol, 1.5 equiv.) was added in three portions over 30 min at 0 °C. The reaction mixture was let warm to rt and stirred for 16 h at rt. Reaction was quenched by addition of NH_4Cl (sat. aq. 25 mL) and extracted with CH_2Cl_2 (3 x 50 mL). Combined organic layers were washed with brine (50 mL) and dried over Na_2SO_4. All volatiles were removed under reduced pressure. Isolated product **7a** was used for next reaction step without further purification.

Nucleoside **7a** (0.66 mmol, 1.0 equiv.) was dissolved in dry 1,4-dioxane (13 mL), an aqueous NH_3-solution (25%, 4 mL) was added. Reaction solution was stirred at rt for 20 h. After full conversion of starting material, reaction was stopped by addition of NH_4Cl (sat. aq. 50 mL). Mixture was extracted with ethyl acetate (3 x 100 mL), combined organic layers were washed with brine (100 mL) and dried over Na_2SO_4. After all volatiles were removed under reduced pressure, crude was purified via flash column chromatography (SiO_2, CH_2Cl_2/MeOH 50:1 → 20:1) to yield **8** as a yellow foam (253 mg, 0.452 mmol, 68 %).



R_f = 0.56 (CH_2Cl_2/MeOH 10:1)

^{1}H-NMR (599 MHz, $CDCl_3$, ppm): δ = 7.72 (s, 1H, C6-H), 7.24 (s, 1H, C4-NH), 6.56 (s, 1H, N3-H), 5.95 (dd, $^3J_{H1'-F2'}$ = 17.4 Hz, $^3J_{H1'-H2'}$= 1.7 Hz, 1H, C1'-H), 5.29 (s, 1H, C-H_m), 4.86 (ddd, $^2J_{H2'-F2'}$ = 52.6 Hz, $^3J_{H2'-H1'}$ = 4.5 Hz, $^3J_{H2'-H3'}$ = 1.7 Hz, 1H, C2'-H), 4.25-4.16 (m, 3H, C_n, C_p, C3'-H)^I 4.06-4.01 (m, 2H, C4'-H, C5'-H) , 3.93-3.87 (m, 2H, C_n, C_p)^I, 3.81-3.77 (m, 1H, C5'-H), 2.18 (dtt, ^2J = 13.7 Hz, ^3J = 2.6 Hz, ^3J = 1.4 Hz, 1H, C_o)^{II}, 1.46 (dtt, ^2J = 13.7 Hz, ^3J = 2.6 Hz, ^3J = 1.4 Hz, 1H, C_o)^{II}, 0.95 (s, 9H, C_{e-g}-H_3)^{III}, 0.89 (s, 9H, C_{h-j} -H_3)^{III}, 0.13 (s, 3H, C_a-H_3)^{IV}, 0.12 (s, 3H, C_b-H_3)^{IV}, 0.10 (s, 3H, C_c-H_3)^{IV} ,0.11 (s, 3H, C_d-H_3)^{IV}.

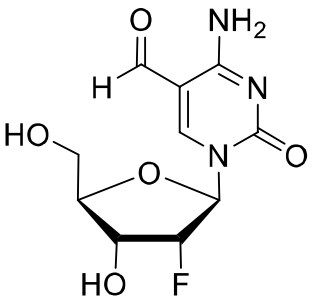
^{13}C-NMR (101 MHz, $CDCl_3$, ppm): δ = 163.2 (C4), 154.5 (C2), 140.5 (C6), 104.3 (C5), 99.3 (C_m) 92.2 (d, $^1J_{C2'-F2'}$ = 192.2 Hz, C2'), 89.7 (d, $^2J_{C1'-F2'}$ = 34.2 Hz, C1'), 83.7 (C4´), 69.0 (d, $^2J_{C3'-F2'}$ = 16.1 Hz, C3'), 67.5 (C_n)^I, 61.5 (C5´), , 26.3 (3C, C_{e-g})^{II}, 25.8 (3C, C_{h-j})^{II}, 25.6 (C_p)^I 18.8 (C_k)^{III}, 18.2 (C_l)^{III}, -4.4 (C_a) ^{VI}, -4.9 (C_b)^{IV}, -5.1 (C_c) ^{VI} -5.1 (C_d)^{IV}.

^{19}F-NMR (377 MHz, $CDCl_3$, ppm): δ = -201.3 (dt, $^2J_{F2'-H2'}$ = 52.6 Hz, $^3J_{F2'-H3',H4'}$ = 17.4 Hz, C2'-F).

HRMS (ESI+): calc. for $C_{25}H_{47}N_3O_6FSi_2^+$ $[M+H]^+$: 560.2982, found: 560.2979.

IR (ATR): ν /cm^{-1} = 2954 (w), 2928 (w), 2891 (w), 2857 (w), 1694 (s), 1463 (m), 1362 (m), 1257 (m), 1127 (m), 1093 (s), 1004 (m), 966 (m), 938 (w), 910 (m), 860 (m), 805 (vs), 777 (vs), 704 (m), 672 (m).

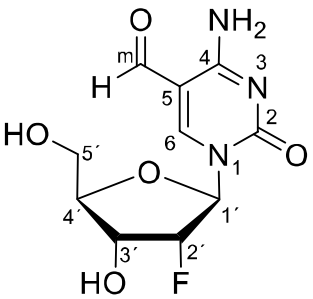
***Synthesis of* 2´-Fluor-5-formyl-cytidine (1)**



1

Nucleoside **8** (480 mg, 0.87 mmol, 1.0 equiv.) and was dissolved in dry THF (30 mL) aq. HCl (1 M, 3.5 mL, 3.45 mmol, 4.0 equiv.) were added. Reaction mixture was stirred at rt for 18 h. Reaction solution was diluted with H_2O (50 mL) and extracted with ethyl acetate (3 x 100 mL). Combined organic layers were washed with brine (50 mL) and dried over Na_2SO_4. All volatiles were removed under reduced pressure. Isolated product **8a** was used for next reaction step without further purification.

In a polypropylene falcon tube Nucleoside **8a** (180 mg, 0.46 mmol, 1.0 equiv.) was dissolved in dry ethyl acetate (7 mL) and cooled to 0 °C. At 0 °C a solution of HF·pyridine (70% HF, 0.18 mL, 6.97 mmol, 15 equiv.) was added, mixture was allowed to warm to rt and was stirred at rt for 20 h. After complete conversion of starting material, reaction was stopped by addition of TMSOMe (2.0 mL, 14.7 mmol, 32 equiv.) at 0 °C, and stirred for 1 h at 0 °C. Subsequently all volatiles were removed under reduced pressure. Crude product was dissolved in ddH_2O and purified by preparative reversed phase HPLC (0% buffer B → 13% buffer B in 45 min; buffer A: ddH2O, buffer B: MeCN). Purified product was lyophilized to yield **1** as a colourless solid (88 mg, 0.32 mmol, 70 %).



R_f = 0.13 (CH_2Cl_2/MeOH 5:1)

^{1}H-NMR (400 MHz, $CDCl_3$, ppm): δ = 9.51 (s, 1H, C_m-H), 8.88 (s, 1H, C6-H), 6.07 (d, $^{3}J_{H1'-F2'}$ = 17.7 Hz, 1H, C1'-H), 5.19 (dd, $^{2}J_{H2'-F2'}$ = 52.4 Hz, $^{3}J_{H2'-H1';H3'}$ = 4.3 Hz, 1H, C2'-H), 4.37 (ddd $^{3}J_{H3'-F2'}$ = 24.1 Hz, $^{3}J_{H3'-H2';H4'}$ = 9.3 Hz, $^{3}J_{H3'-H2';H4'}$ = 4.3 Hz, 1H, C3'-H), 4.27-4.19 (m, 1H, C4'-H) , 4.13 (dd, $^{2}J_{H5'a-H5'b}$ = 13.2 Hz, $^{3}J_{H5'a-H4'}$ = 2.3 Hz, 1H, C5'-H_a)[l], 3.91 (dd, $^{2}J_{H5'b-H5'a}$ = 13.2 Hz, $^{3}J_{H5'b-H4'}$ = 3.5 Hz, 1H, $C_{5'}$-H_b)[l].

^{13}C-NMR (800 MHz, $CDCl_3$, ppm): δ = 190.2 (C_m), 162.8 (C4), 155.0 (C2), 154.9 (C6), 105.9 (C5), 93.7 (d, $^{1}J_{C2'-F2'}$ = 185.2 Hz, C2'), 89.9 (d, $^{2}J_{C1'-F2'}$ = 34.9 Hz, C1'), 82.4 (C4´), 67.2 (d, $^{2}J_{C3'-F2'}$ = 16.8 Hz, C3'), 59.1 (C5).

^{19}F-NMR (400 MHz, $CDCl_3$, ppm): δ = -201.6 (ddd, $^{2}J_{F2'-H2'}$ = 52.4 Hz, $^{3}J_{F2'-H3'}$ = 24.3 Hz, $^{3}J_{F2'-H1'}$ = 17.8 Hz C2'-F).

HRMS (ESI+): calc. for $C_{10}H_{13}N_3O_5F^+$ $[M+H]^+$: 274.0834, found: 274.0835.

Supplementary tables

Table SI-1[2]: Compound-dependent LC-MS/MS-parameters used for the analysis of genomic DNA fed with fluorinated nucleosides. *CE: collision energy, CAV: collision cell accelerator voltage, EMV: electron multiplier voltage. The nucleosides were analysed in the positive ($[M+H]^+$ species) as well as the negative ($[M-H]^-$ species) ion selected reaction monitoring mode (SRM).*

compound	Precursor ion (m/z)	MS1 Resolution	Product Ion (m/z)	MS2 Resolution	Dwell time [ms]	CE (V)	CAV (V)	Polarity
Time segment 1.5-3.0 min								
F-dC	246.09	Wide	112.06	Wide	70	15	3	Positive
[$^{15}N_2$]-F-dC	248.08	Wide	114.04	Wide	70	15	3	Positive
[$^{15}N_2$]-cadC	274.08	Wide	158.03	Wide	40	5	5	Positive
cadC	272.09	Wide	156.04	Wide	40	5	5	Positive
F-hmdC	276.10	Wide	142.06	Wide	50	10	3	Positive
[$^{15}N_2$]-F-hmdC	278.09	Wide	144.06	Wide	50	10	3	Positive
[$^{15}N_2$,D_2]-hmdC	262.12	Wide	146.07	Wide	25	27	1	Positive
hmdC	258.11	Wide	142.06	Wide	25	27	1	Positive
[D_3]-mdC	245.13	Wide	129.09	Wide	50	60	1	Positive
mdC	242.11	Wide	126.07	Wide	50	60	1	Positive
Time segment 3.0-4.7 min								
F-hmdU	275.07	Wide	255.06	Wide	80	3	7	Negative
F-dU	245.06	Wide	225.06	Wide	80	3	5	Negative
[D_3]-F-mdC	263.12	Wide	129.09	Wide	80	15	3	Positive
F-mdC	260.10	Wide	126.07	Wide	80	15	3	Positive
F-cadC	290.08	Wide	156.04	Wide	80	5	5	Positive
Time segment 4.7-10 min								
F-fdC	274.08	Wide	140.05	Wide	90	15	3	Positive
[$^{15}N_2$]-F-fdC	276.08	Wide	142.04	Wide	90	15	3	Positive
F-fdU	273.05	Wide	253.05	Wide	30	3	5	Negative

F-dT	259.07	Wide	239.07	Wide	70	3	5	Negative
dT	243.10	Wide	127.05	Wide	20	5	5	Negative
[$^{15}N_2$]-fdC	258.09	Wide	142.04	Wide	30	5	5	Positive
fdC	256.09	Wide	140.05	Wide	30	5	5	Positive
[$^{15}N_2$]-fdC	142.04	Wide	98.04	Wide	20	13	7	Positive
fdC	140.05	Wide	97.04	Wide	20	13	7	Positive

Table SI-2: Compound-dependent LC-MS/MS ranges of the corresponding linear equations.

compound	n (ULOQ)	n (LLOQ)	A/A* (ULOQ)	A/A* (LLOQ)
F-fdC	25.1 fmol	0.390 fmol	0.2267	0.003345
F-dC	800 fmol	3.13 fmol	4.074	0.01512
F-mdC	49.9 fmol	0.780 fmol	0.04174	0.0005833

Table SI-3: Compound-dependent LC-MS/MS-parameters used for the analysis of genomic DNA fed with fluorinated nucleosides and isotopically labelled dT. *CE: collision energy, CAV: collision cell accelerator voltage, EMV: electron multiplier voltage. The nucleosides were analysed in the positive ($[M+H)]^+$ species) as well as the negative ($[M-H]^-$ species) ion selected reaction monitoring mode (SRM).*

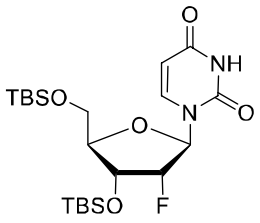
compound	Precursor ion (m/z)	MS1 Resolution	Product Ion (m/z)	MS2 Resolution	Dwell time [ms]	CE (V)	CAV (V)	Polarity
Time segment 1.5-3.7 min								
F-dC	246.09	Wide	112.06	Wide	70	15	3	Positive
[$^{15}N_2$]-F-dC	248.08	Wide	114.04	Wide	70	15	3	Positive
[$^{15}N_2$]-cadC	274.08	Wide	158.03	Wide	40	5	5	Positive
cadC	272.09	Wide	156.04	Wide	40	5	5	Positive
F-hmdC	276.10	Wide	142.06	Wide	50	10	3	Positive
[$^{15}N_2$]-F-hmdC	278.09	Wide	144.06	Wide	50	10	3	Positive
[$^{15}N_2$,D_2]-hmdC	262.12	Wide	146.07	Wide	25	27	1	Positive
hmdC	258.11	Wide	142.06	Wide	25	27	1	Positive
[D_3]-mdC	245.13	Wide	129.09	Wide	50	60	1	Positive
mdC	242.11	Wide	126.07	Wide	50	60	1	Positive
Time segment 3.7-4.7 min								

F-hmdU	275.07	Wide	255.06	Wide	80	3	7	Negative
F-dU	245.06	Wide	225.06	Wide	80	3	5	Negative
$[D_3]$-F-mdC	263.12	Wide	129.09	Wide	80	15	3	Positive
F-mdC	260.10	Wide	126.07	Wide	80	15	3	Positive
F-cadC	290.08	Wide	156.04	Wide	80	5	5	Positive
Time segment 4.7-12 min								
F-fdC	274.08	Wide	140.05	Wide	90	15	3	Positive
$[^{15}N_2]$-F-fdC	276.08	Wide	142.04	Wide	90	15	3	Positive
F-fdU	273.05	Wide	253.05	Wide	30	3	5	Negative
F-dT	259.07	Wide	239.07	Wide	70	3	5	Negative
dT	243.10	Wide	127.05	Wide	20	5	5	Negative
$[^{15}N_2]$-fdC	258.09	Wide	142.04	Wide	30	5	5	Positive
fdC	256.09	Wide	140.05	Wide	30	5	5	Positive
$[^{15}N_2]$-fdC	142.04	Wide	98.04	Wide	20	13	7	Positive
fdC	140.05	Wide	97.04	Wide	20	13	7	Positive
$[^{13}C_5{}^{15}N_2]$-dT	250.12	Wide	129.04	Wide	20	3	5	Positive
$[^{13}C_{10}{}^{15}N_2]$-dT	255.12	Wide	134.06	Wide	20	3	5	Positive

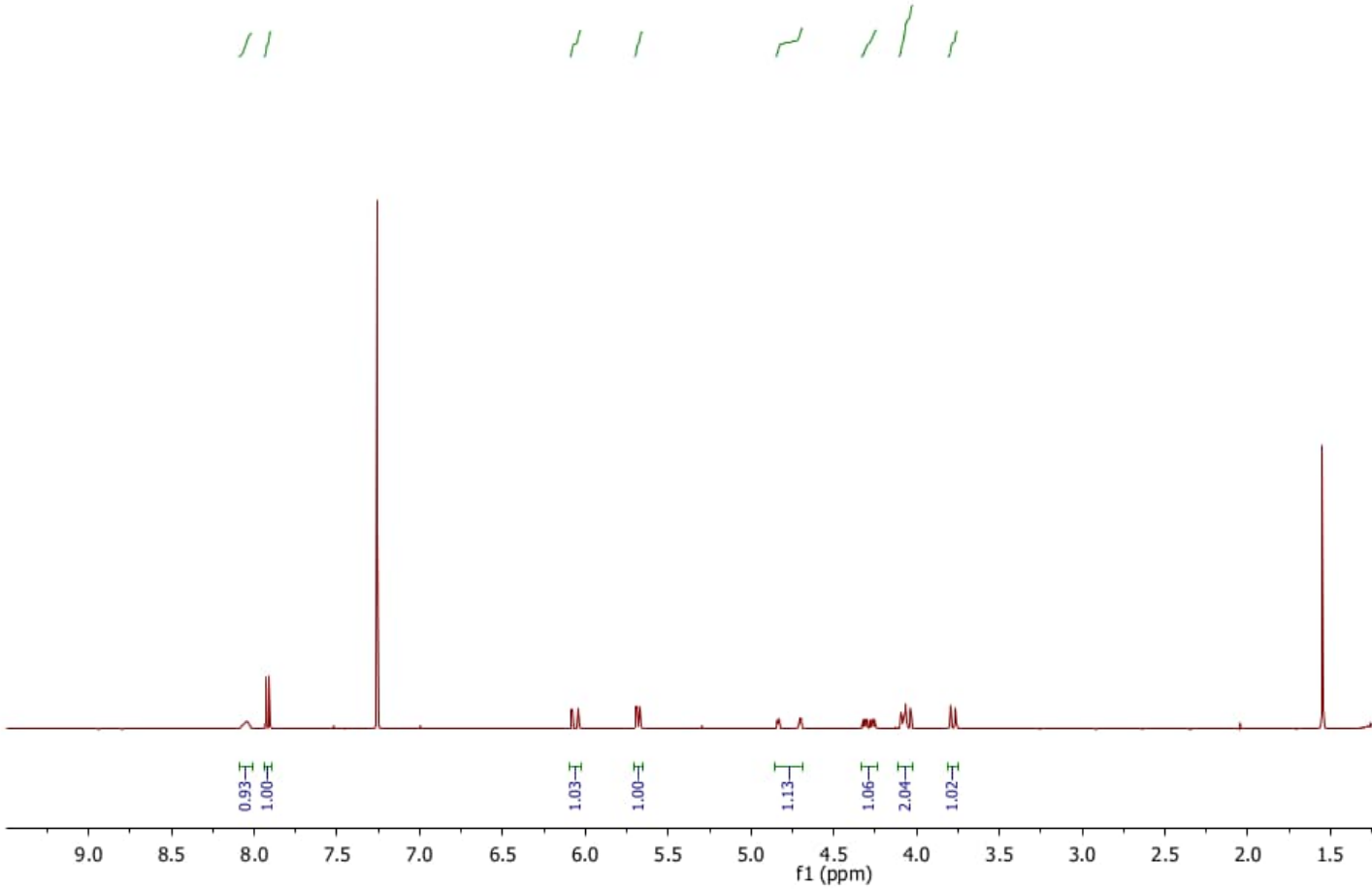
Appendix

NMR Spectra

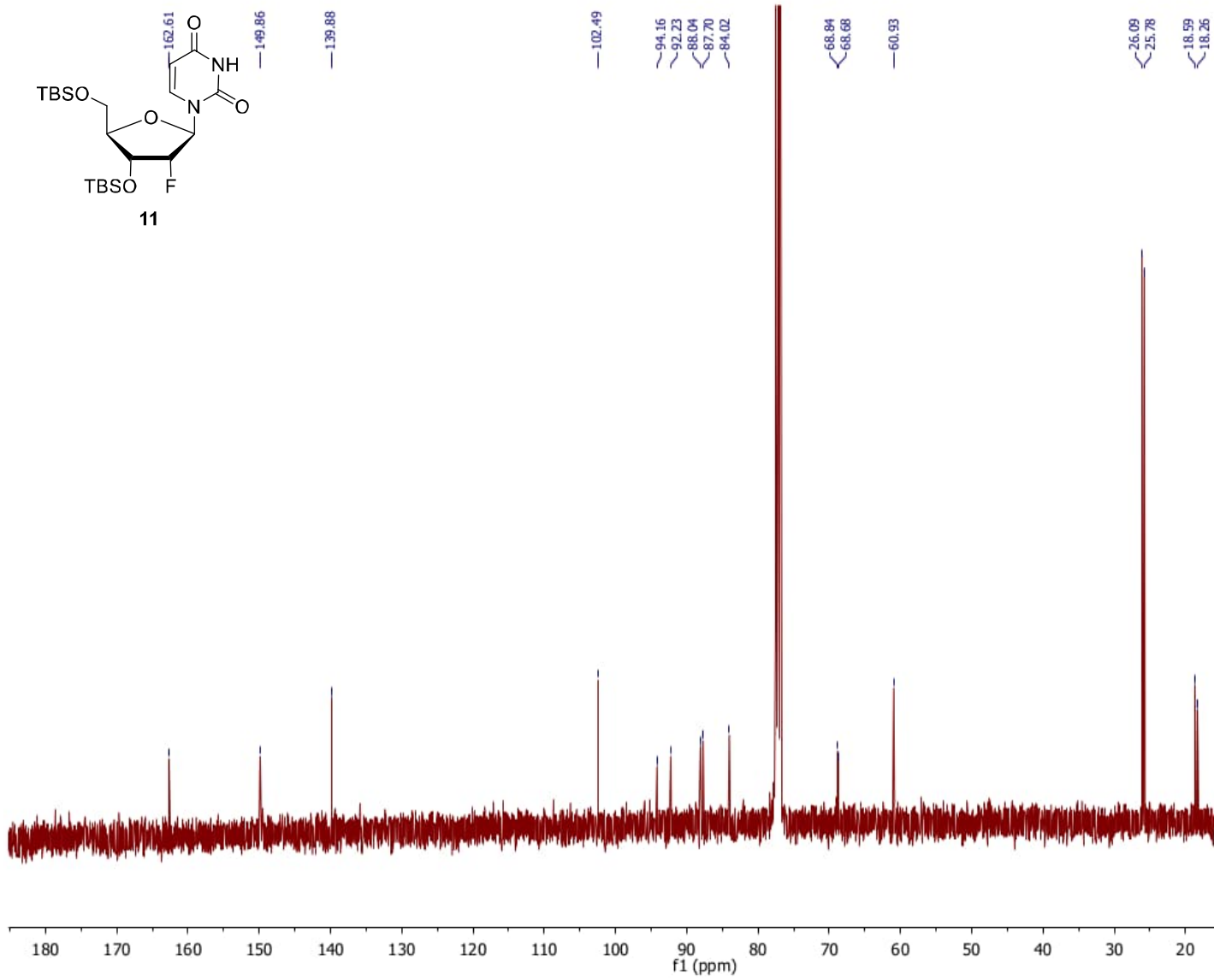
Compound 11, ^{1}H-NMR, 400 MHz, $CDCl_3$



11

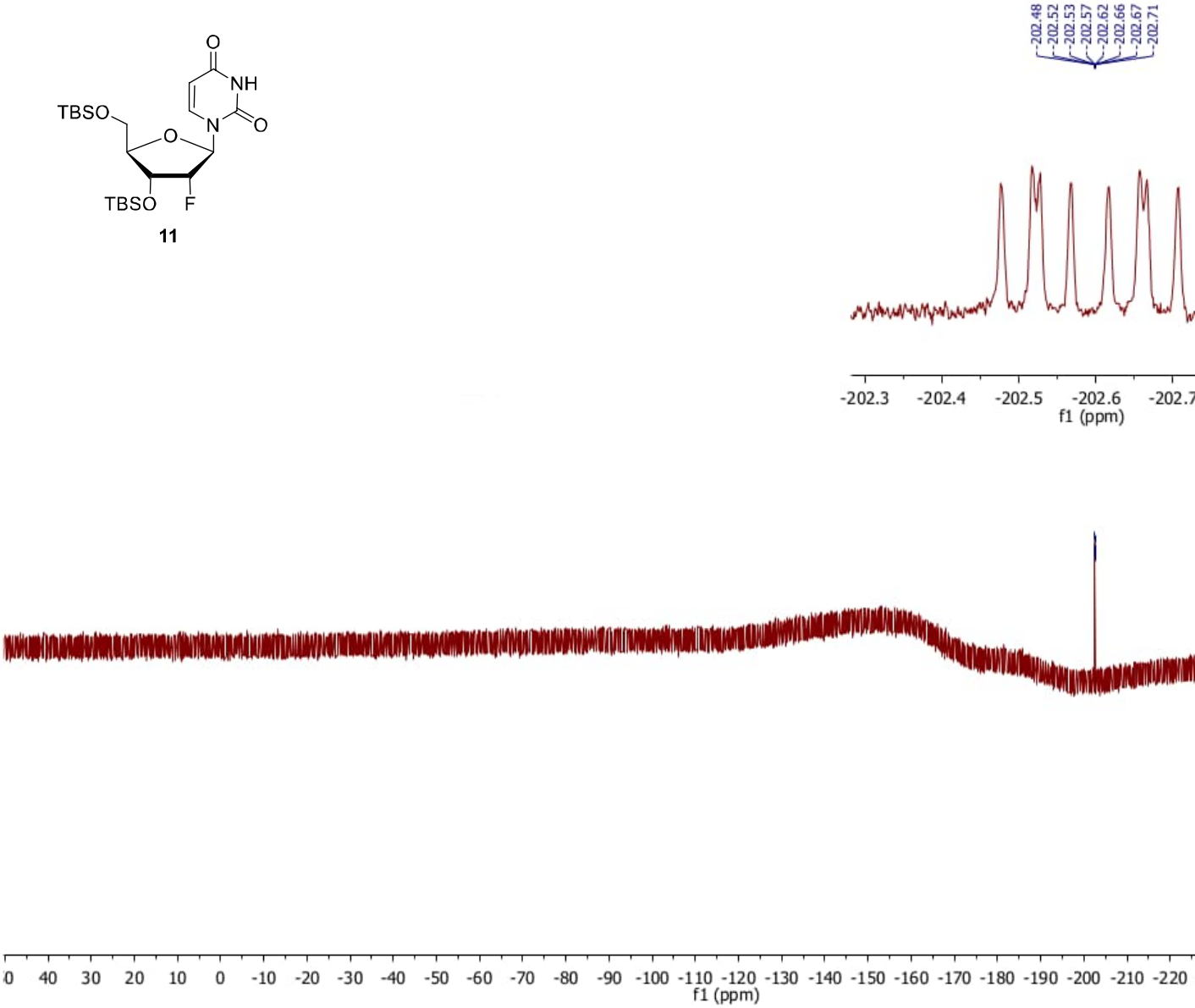


Compound 11, ^{13}C-NMR, 400 MHz, $CDCl_3$

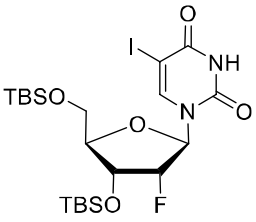


Compound 11, ^{19}F-NMR, 377 MHz, $CDCl_3$

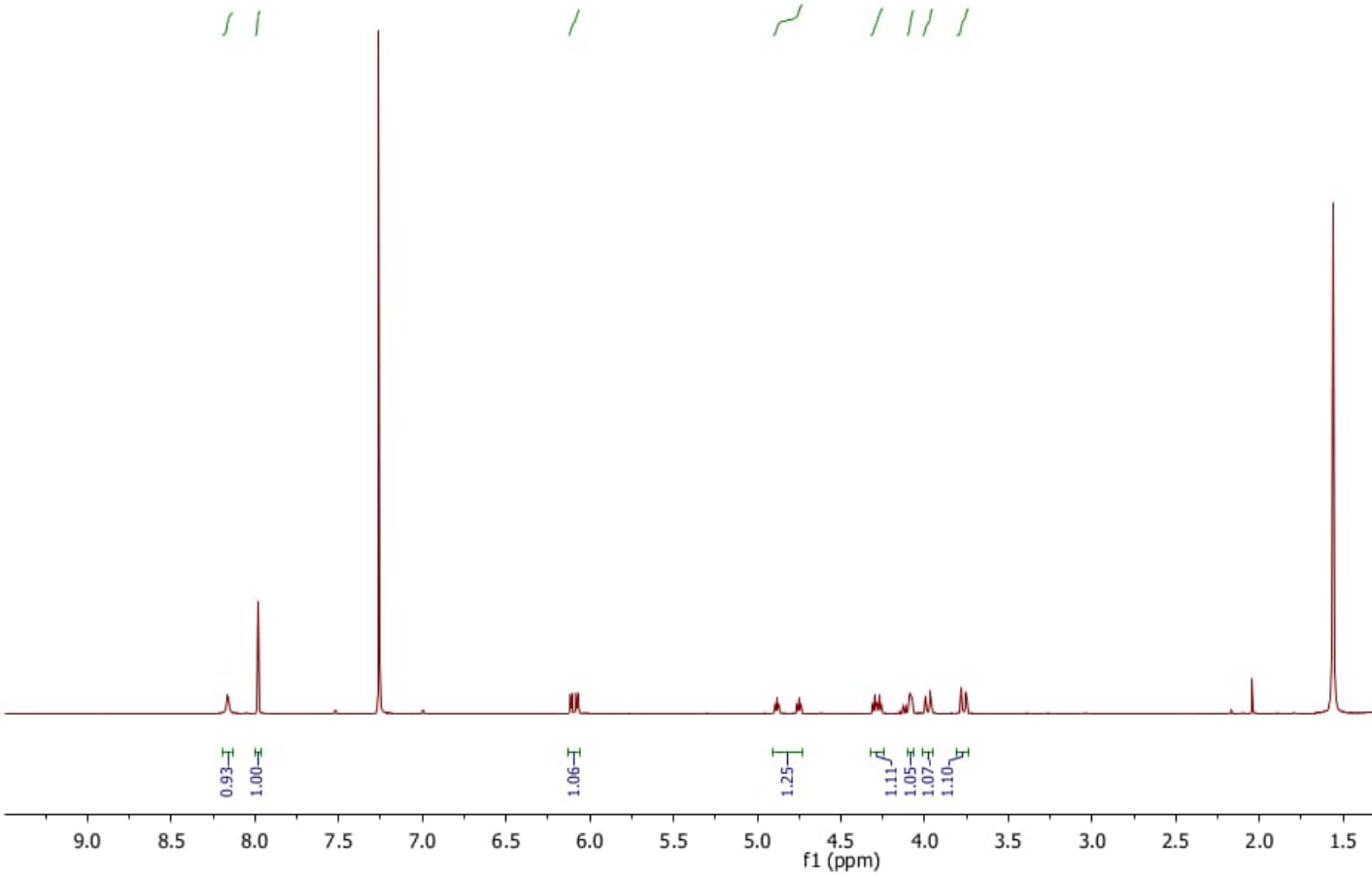
11



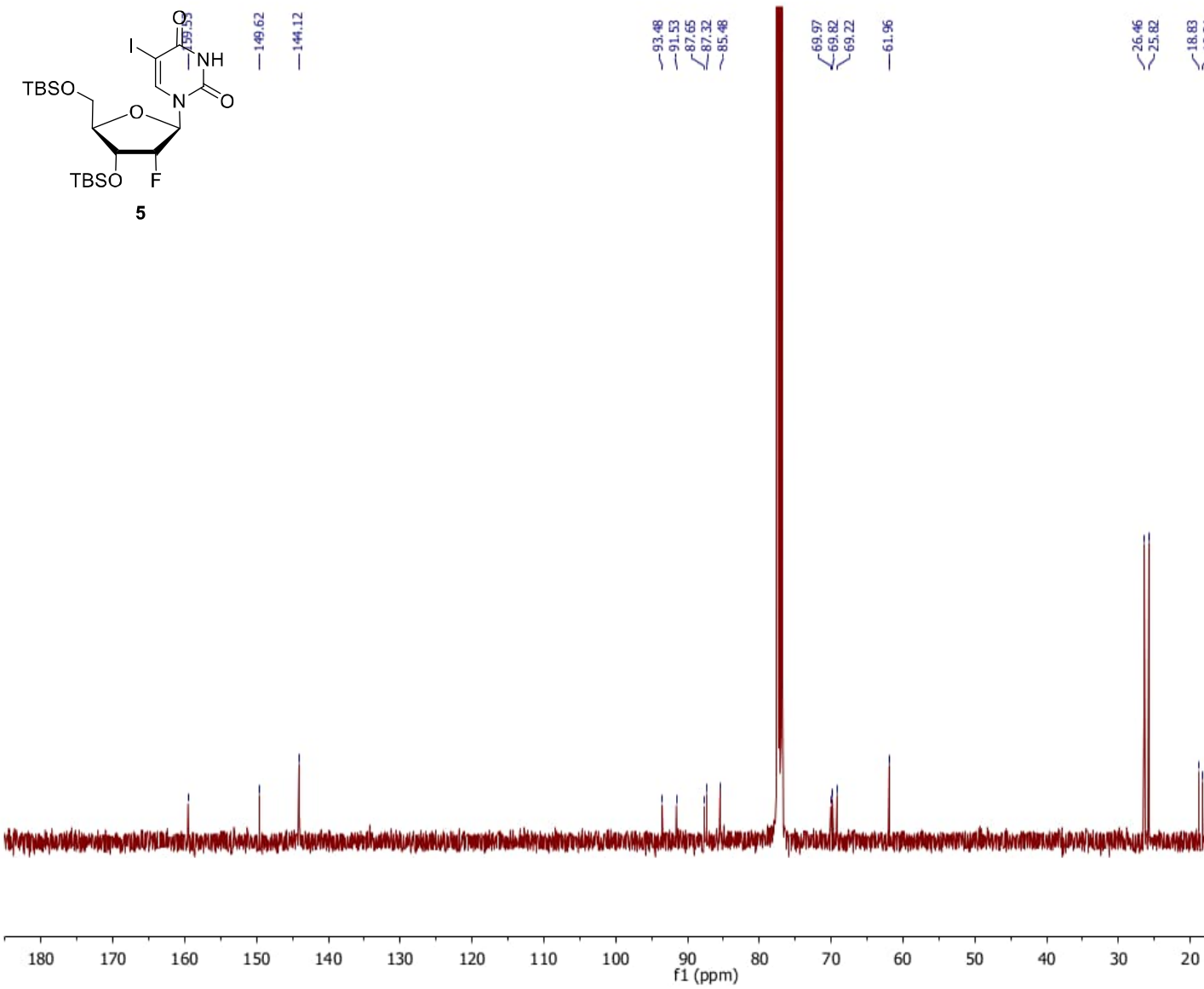
Compound 5, ^{1}H-NMR, 400 MHz, $CDCl_3$



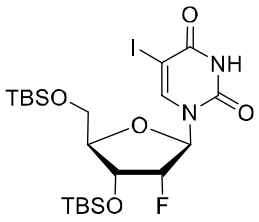
5



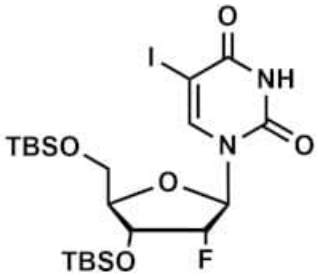
Compound 5, ^{13}C-NMR, 4800 MHz, $CDCl_3$

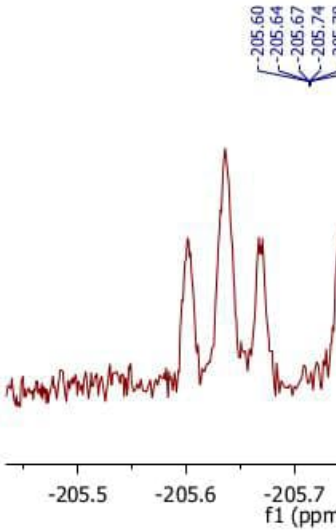


Compound 5, ^{19}F-NMR, 377 MHz, $CDCl_3$



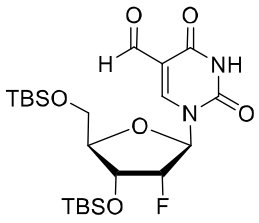
5



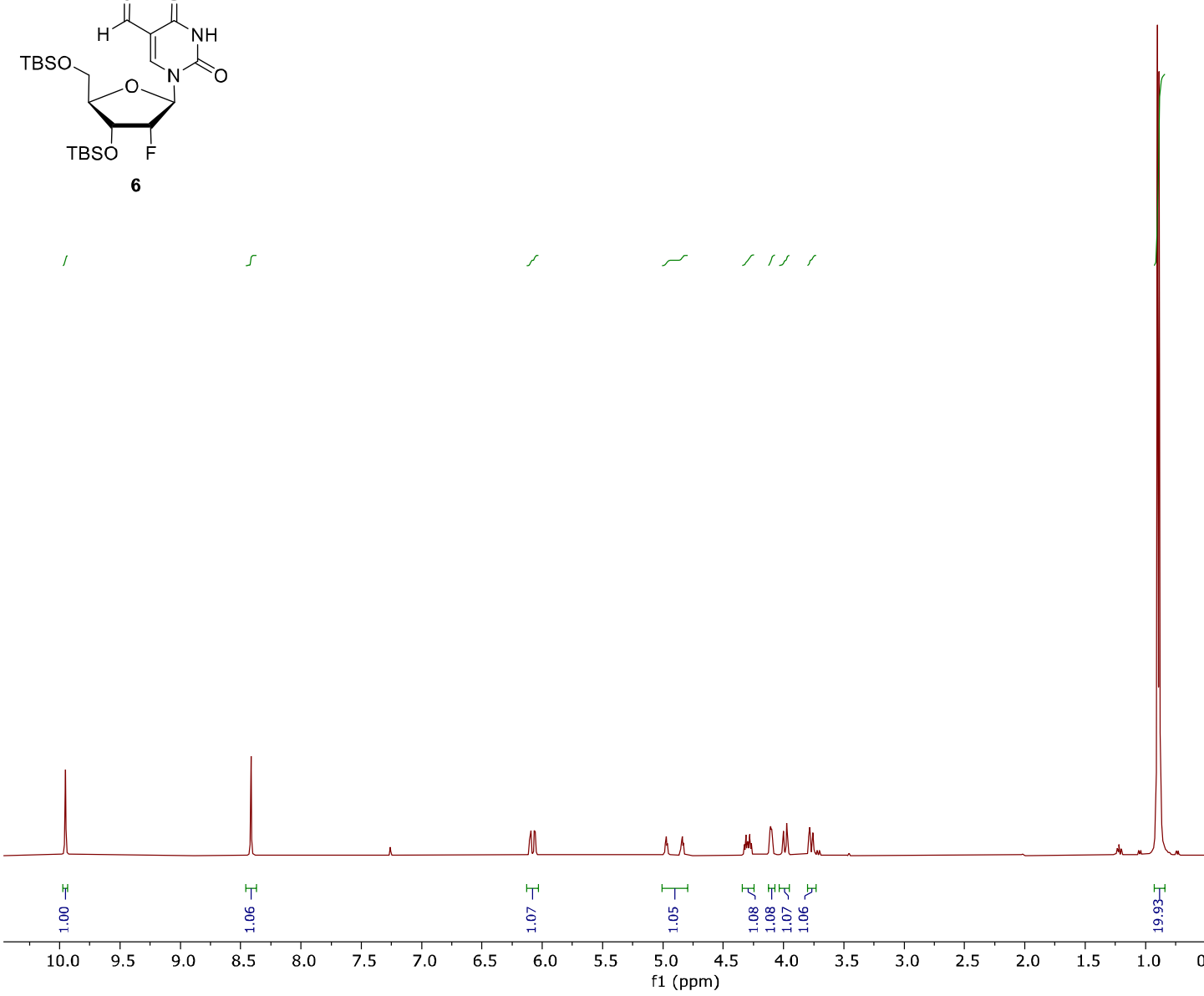




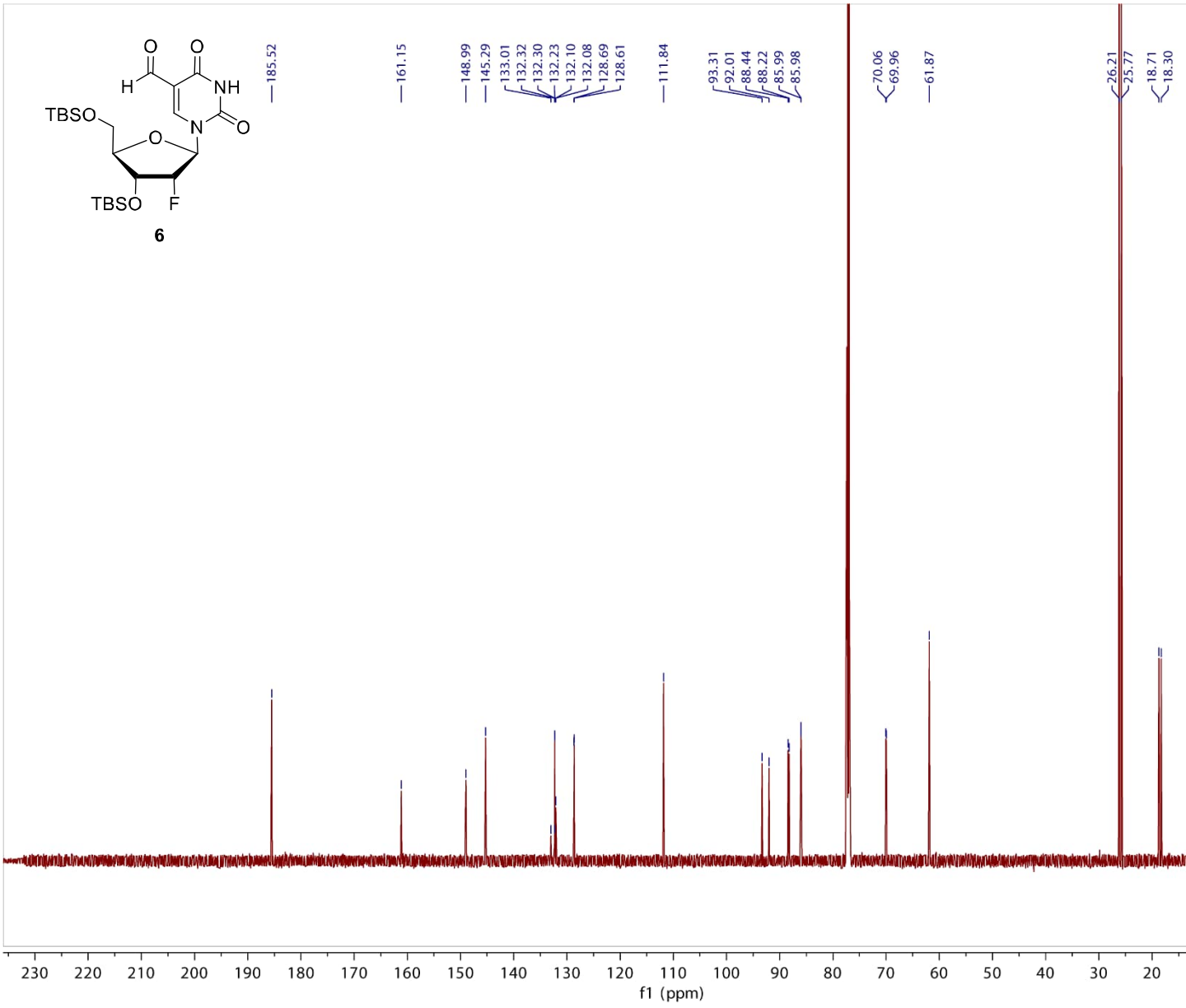
Compound 6, ^{1}H-NMR, 400 MHz, $CDCl_3$



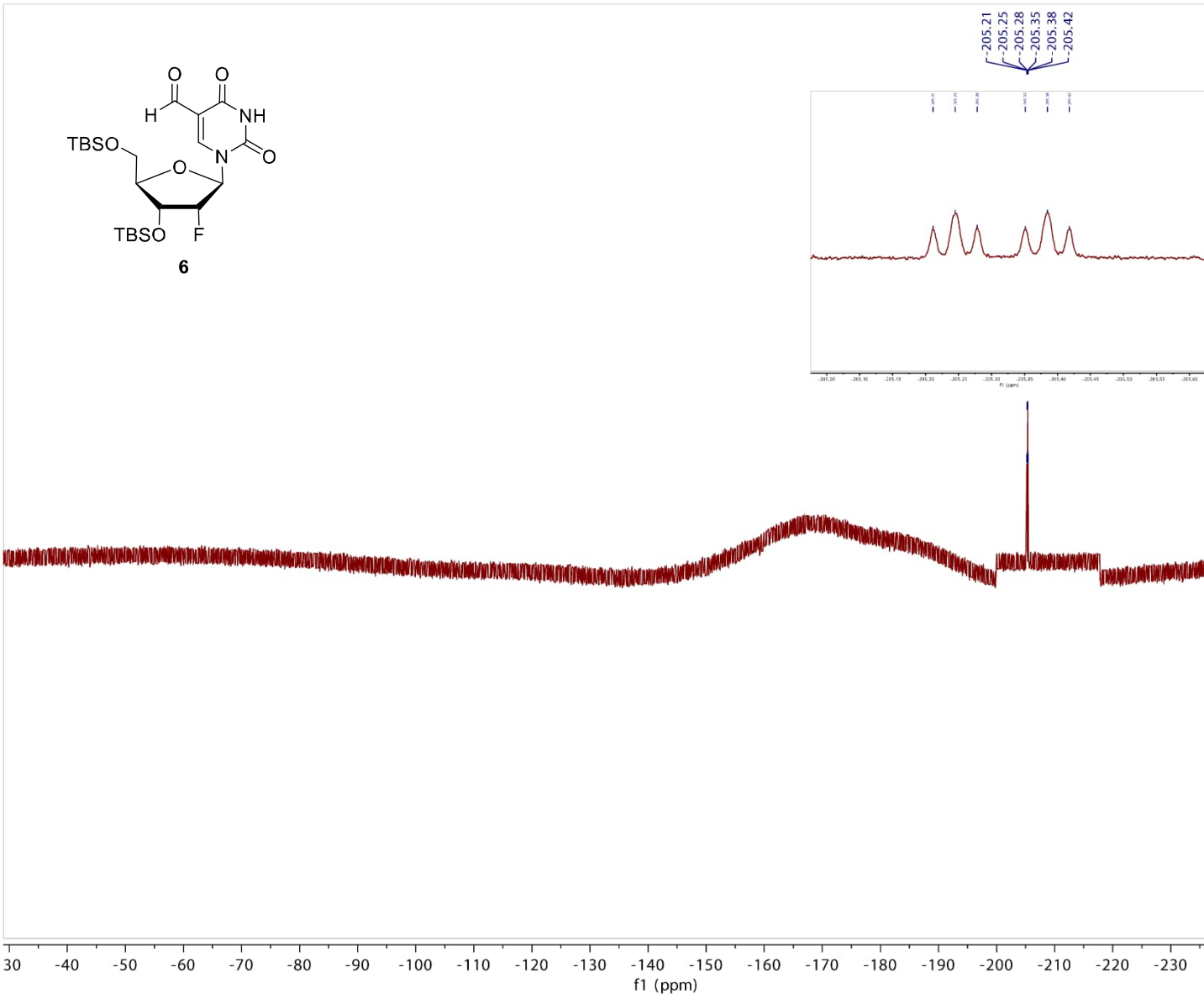
6



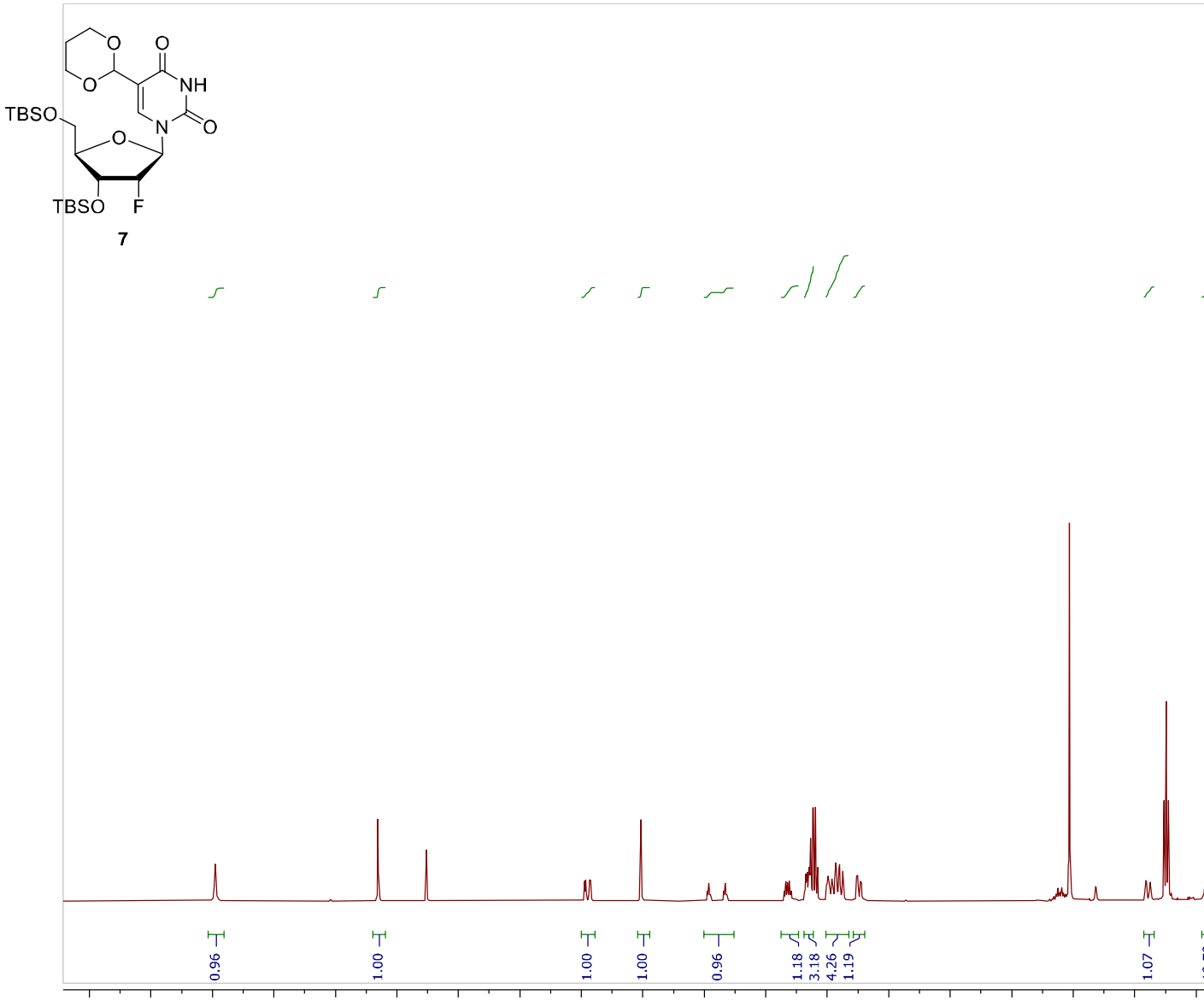
Compound 6, ^{13}C-NMR, 400 MHz, $CDCl_3$



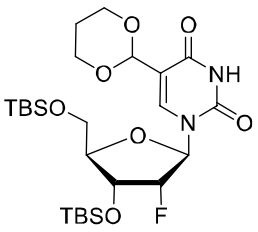
Compound 6, ^{19}F-NMR, 377 MHz, $CDCl_3$



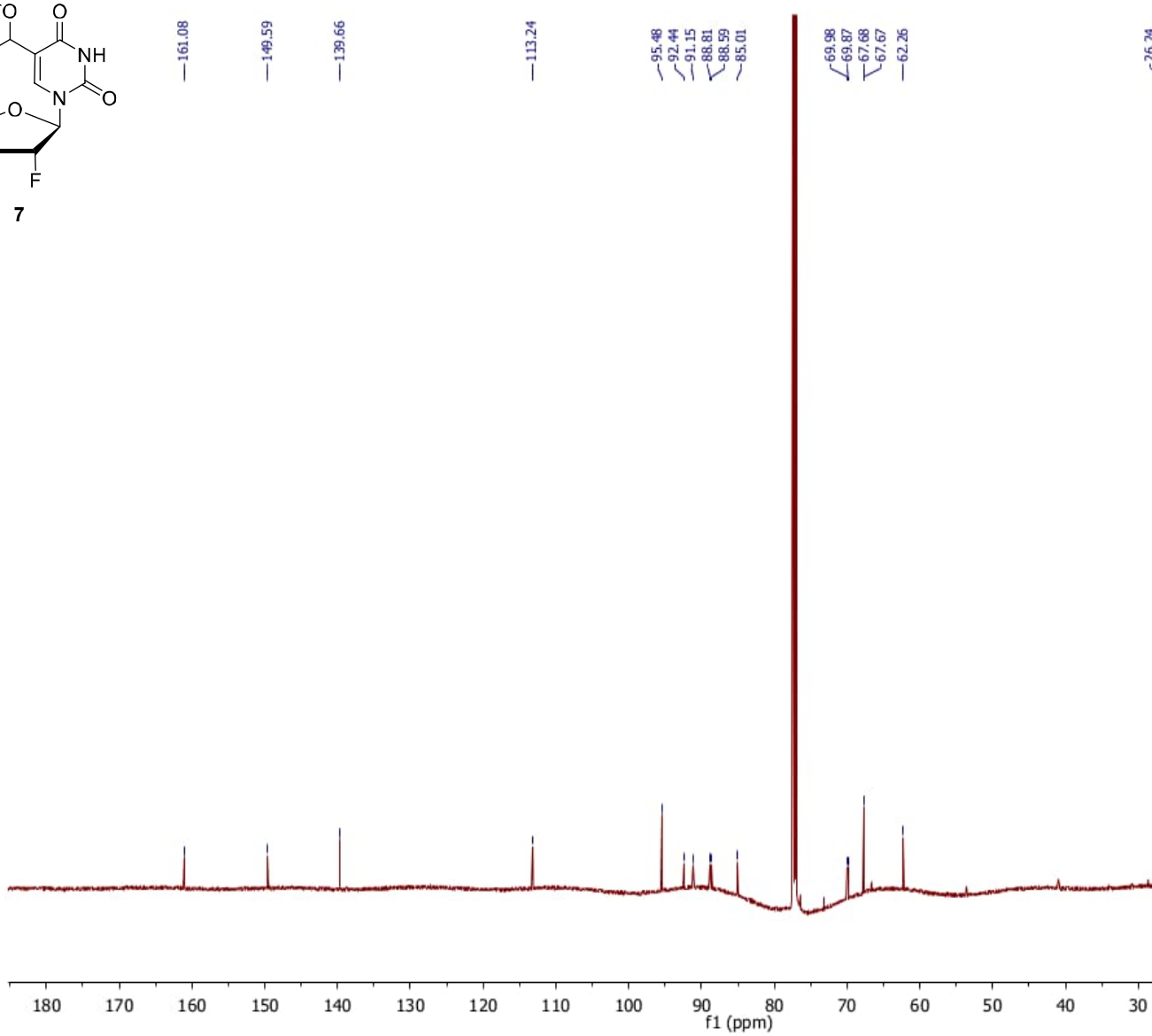
Compound 7, ^{1}H-NMR, 400 MHz, $CDCl_3$



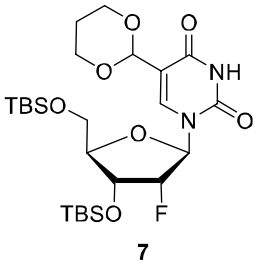
Compound 7, ^{13}C-NMR, 400 MHz, $CDCl_3$



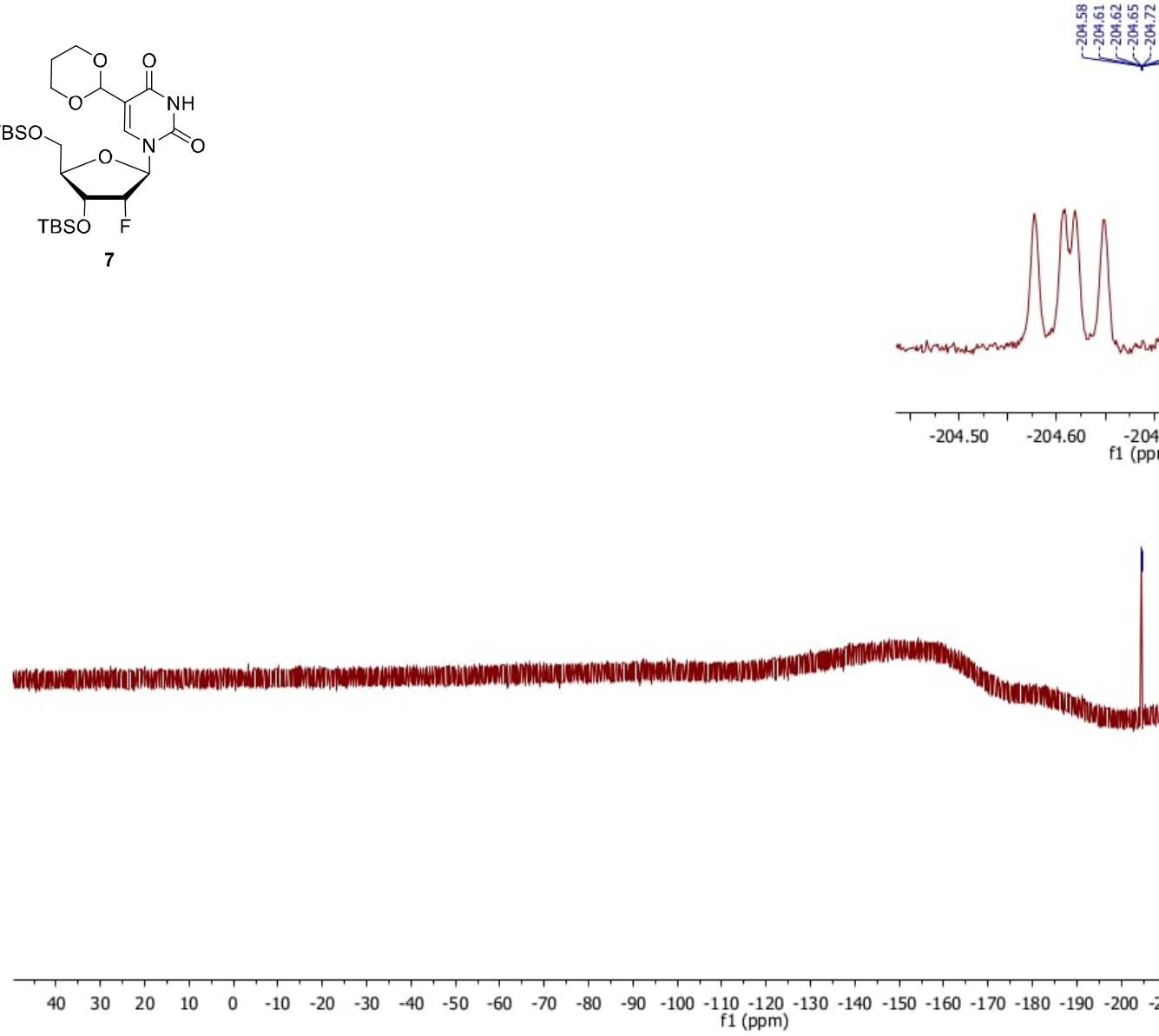
7



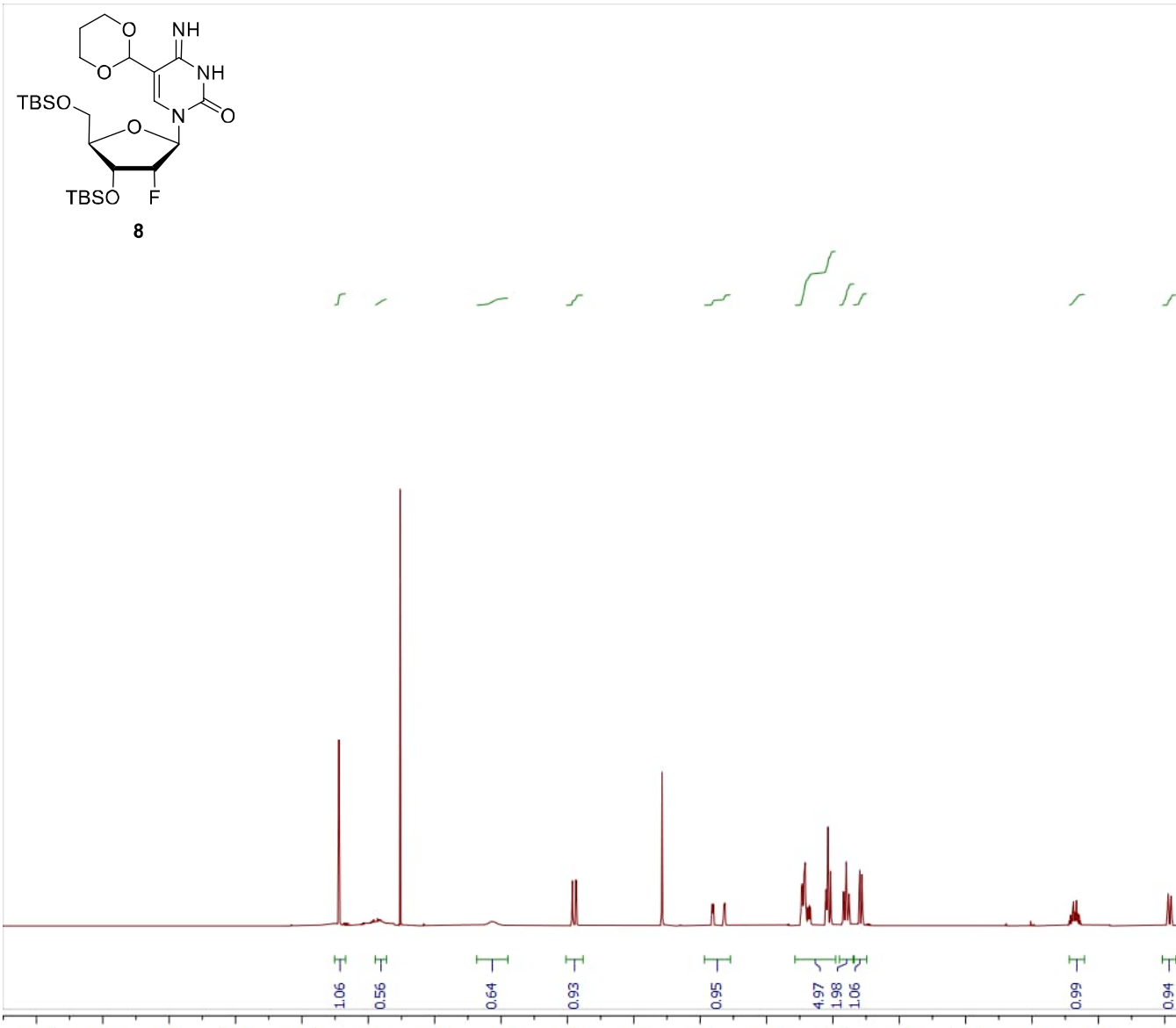
Compound 7, ^{19}F-NMR, 377 MHz, $CDCl_3$



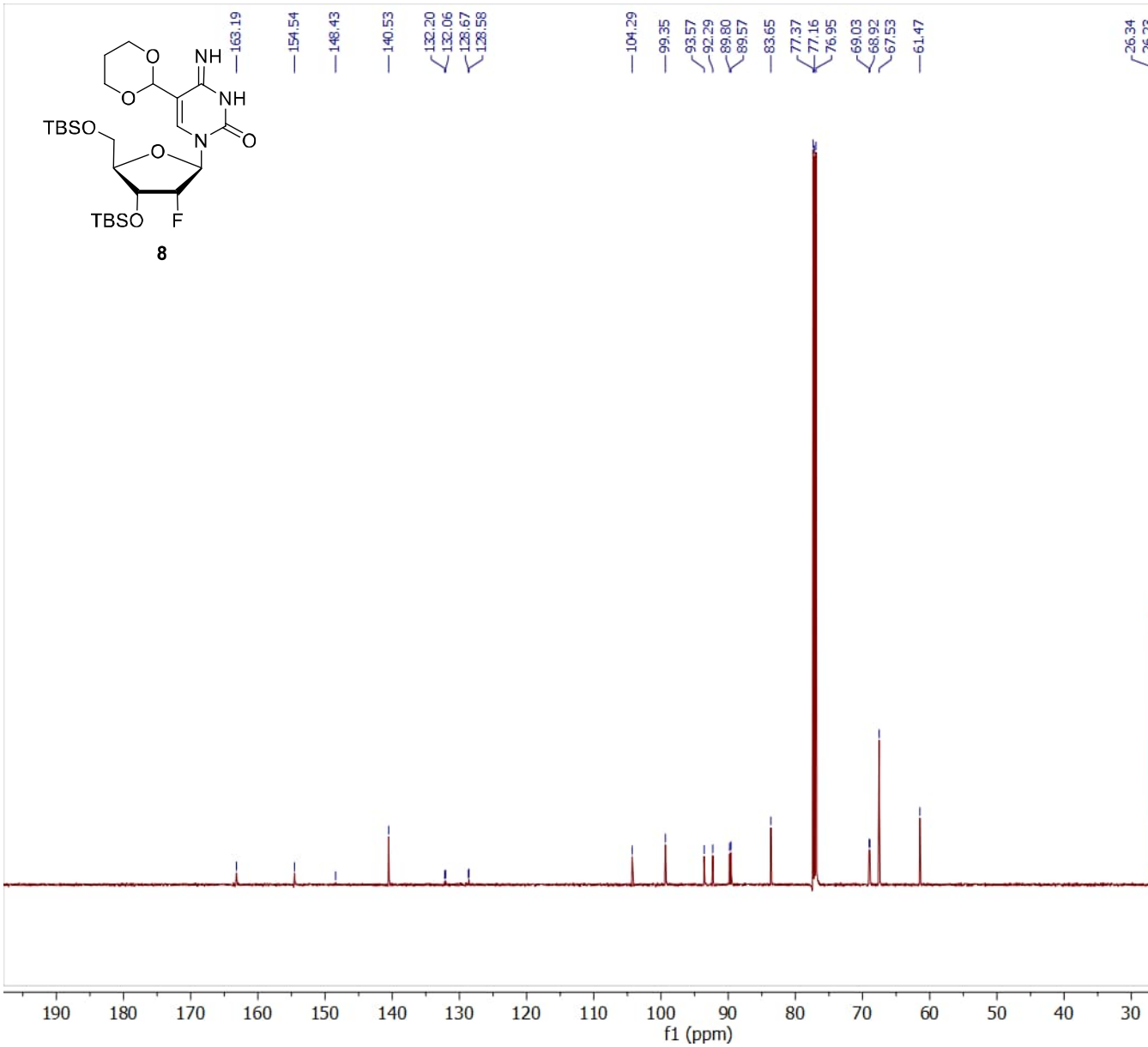
7



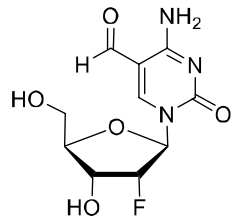
Compound 8, ^{1}H-NMR, 400 MHz, $CDCl_3$



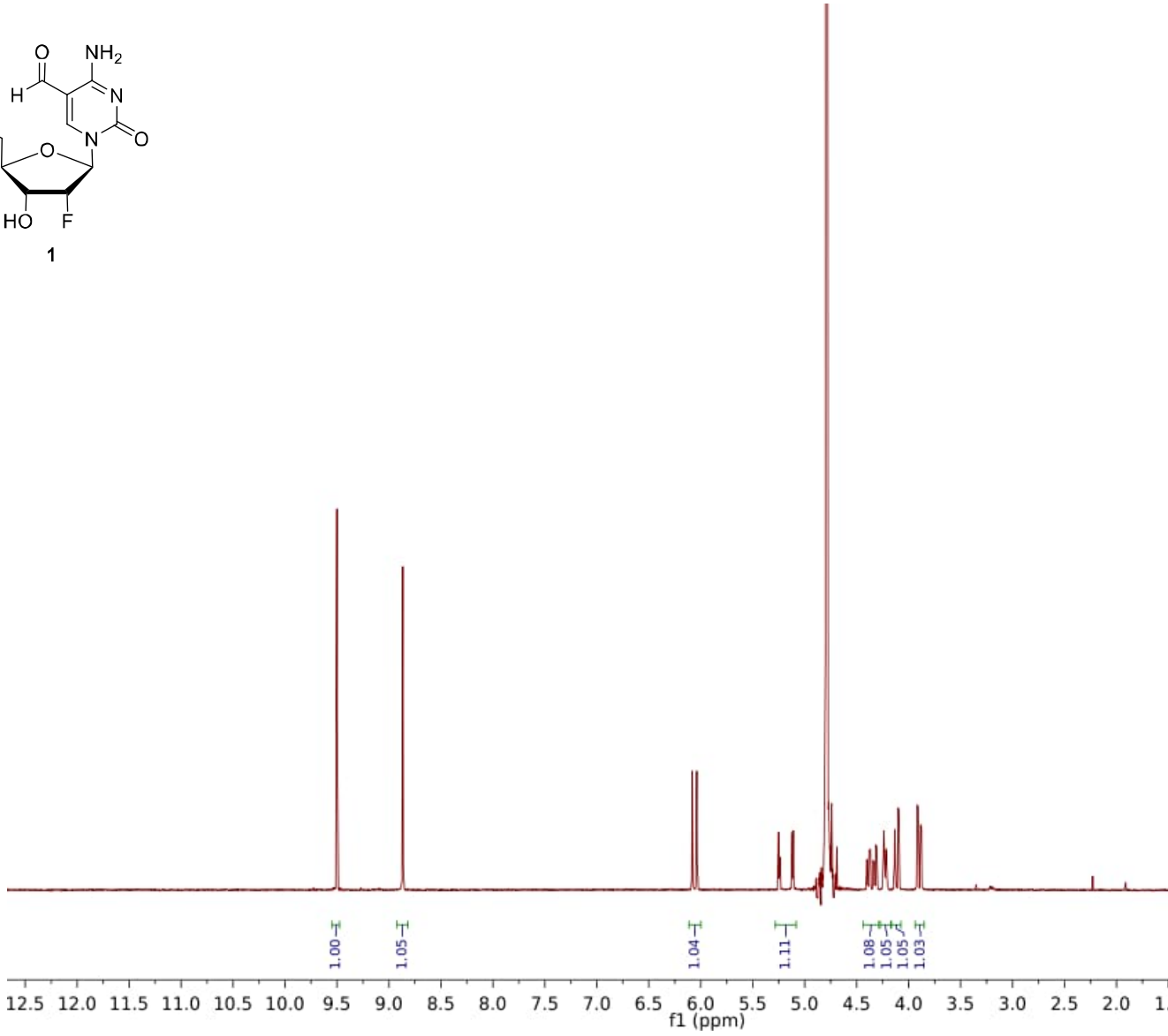
Compound 7, ^{13}C-NMR, 600 MHz, $CDCl_3$



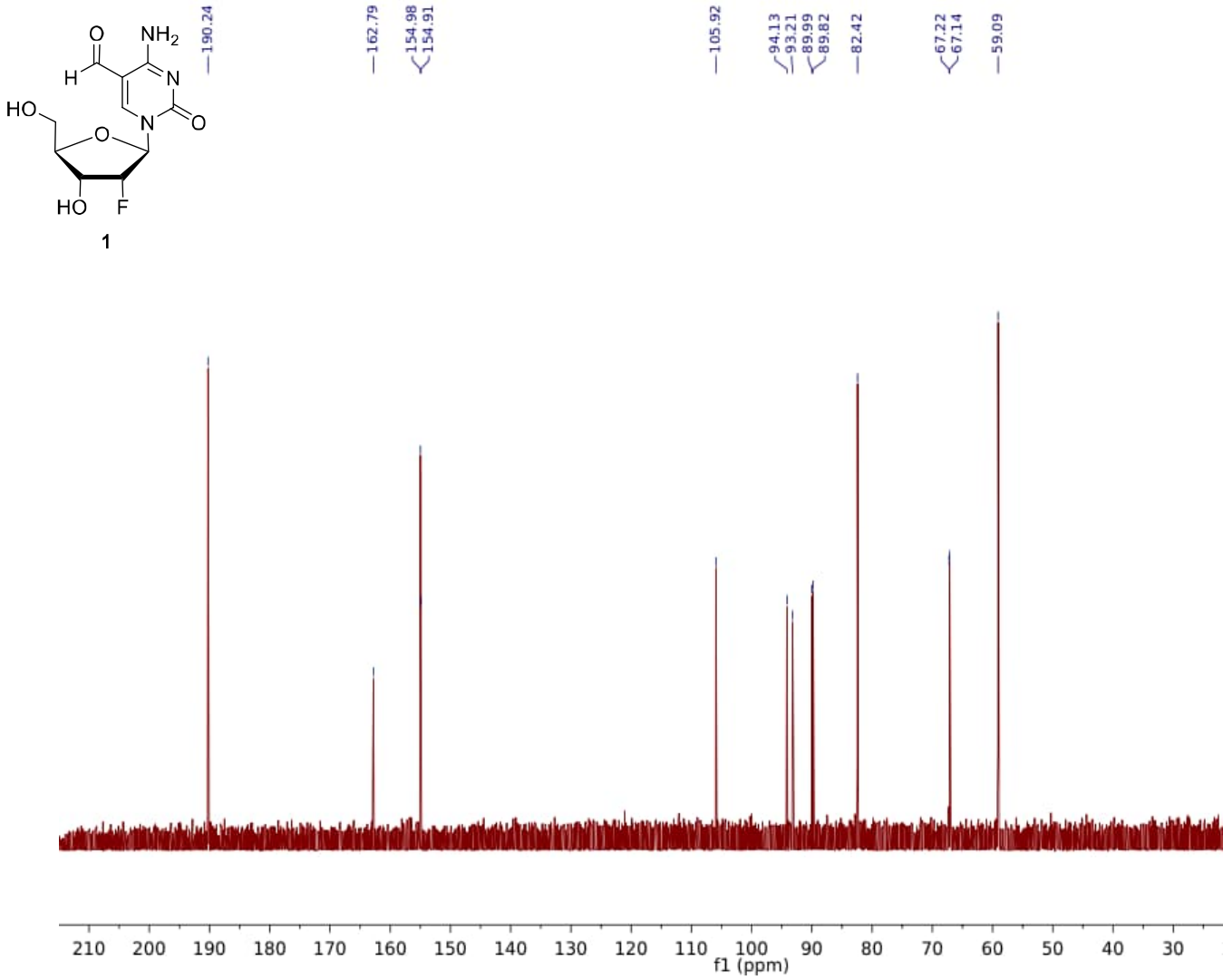
Compound 1, ^{1}H-NMR, 400 MHz, D_2O



1

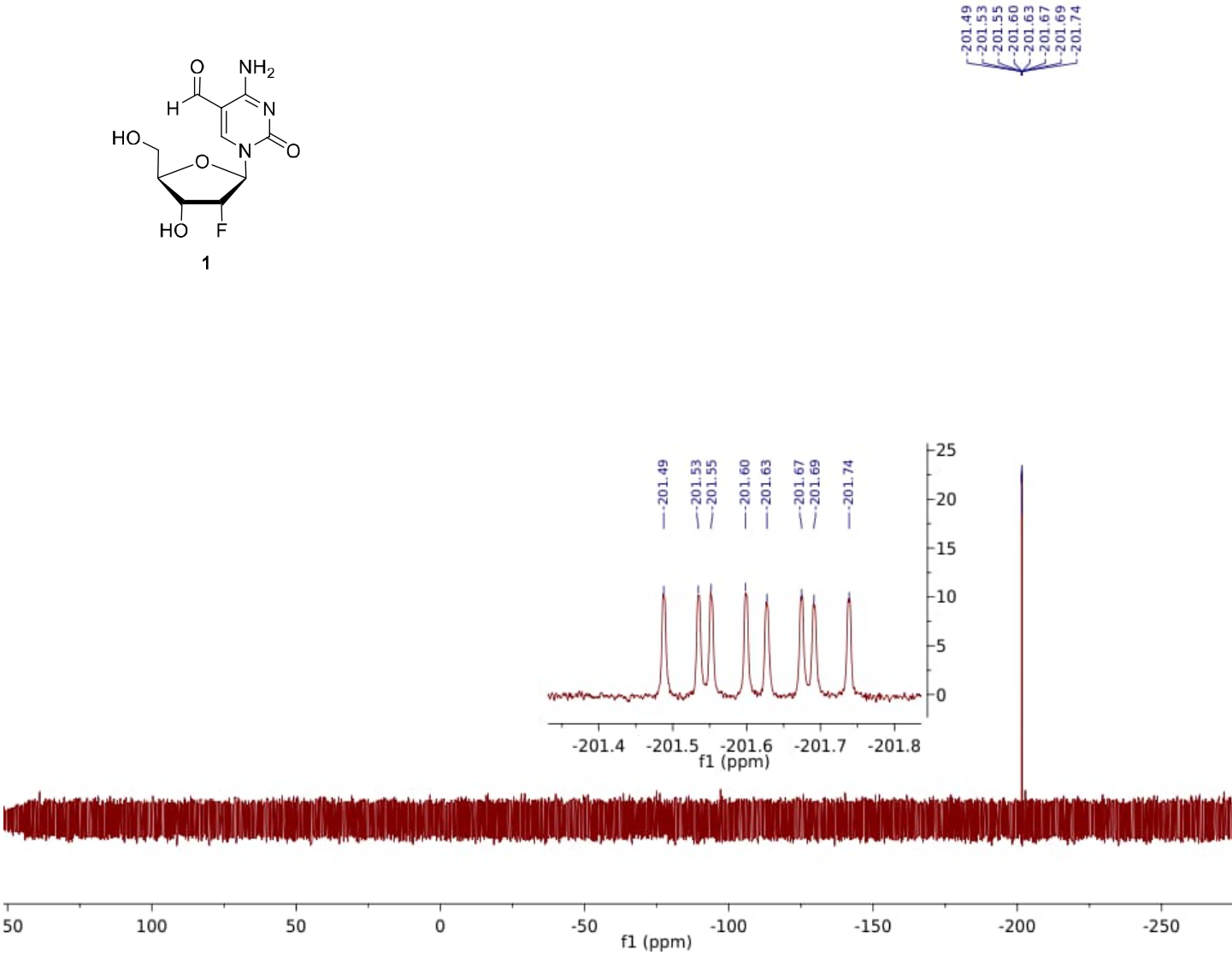


Compound 1, ^{13}C-NMR, 800 MHz, D_2O



Compound 1, ^{19}F-NMR, 400 MHz, D_2O

1



References

Uncategorized References

[1] W. J. Lowary PT, *J Mol. Biol.* **1998**, *276*, 19-42.
[2] K. Iwan, R. Rahimoff, A. Kirchner, F. Spada, A. S. Schröder, O. Kosmatchev, S. Ferizaj, J. Steinbacher, E. Parsa, M. Müller, T. Carell, *Nat. Chem. Biol.* **2018**, *14*, 7
[3] V. Busskamp, N. E. Lewis, P. Guye, A. H. Ng, S. L. Shipman, S. M. Byrne, N. E. Sanjana, J. Murn, Y. Li, S. Li, M. Stadler, R. Weiss, G. M. Church, *Mol. Syst. Biol.*
[4] F. R. Traube, S. Schiffers, K. Iwan, S. Kellner, F. Spada, M. Müller, T. Carell, *Nat. Protoc.* **2019**, *14*, 283-312.
[5] A. S. Schröder, O. Kotljarova, E. Parsa, K. Iwan, N. Raddaoui, T. Carell, *Org. Lett.* **2016**, *18*, 4368-4371.

8.2 Supporting Information: Intragenomic decarboxylation of 5-carboxycytidine

E. Kaminska*, E. Korytiaková*, A. Reichl, M. Müller, T. Carell *Angew. Chem. Int. Ed.* **2021,** *60,* 23207-23211.

*: Diese Autoren haben zu gleichen Teilen beigetragen

Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

Intragenomic Decarboxylation of 5-Carboxy-2′-deoxycytidine

Ewelina Kamińska[+], *Eva Korytiaková*[+], *Andreas Reichl, Markus Müller, and Thomas Carell**

anie_202109995_sm_miscellaneous_information.pdf

Table of Contents

Supplementary figures 3
Methods and Materials 9
Chemical Synthesis 9
Cell culture 10
Cell Harvesting and Isolation of genomic DNA 11
DNA digestion 11
LC/MS-MS analysis of DNA samples[4] 12
Quantification of nucleosides in DNA samples 12
Soluble nucleoside pool extraction and purification 12
Synthetic procedures 14
3',5'-di-*O*-Acetyl-2'-deoxy-2'-(*R*)-fluorouridine (9) 15
3',5'-di-*O*-Acetyl-2'-deoxy-2'-(*R*)-fluoro-3-nitrouridine (10) 16
3',5'-di-*O*-Acetyl-2'-deoxy-2'-(*R*)-fluoro-(N^3-^{15}N)-uridine (11) 17
3',5'-bis-*O*-(tert-Butyl(dimethyl)silyl)-2'-deoxy-2'-(*R*)-fluoro-(N^3-^{15}N)-uridine (12) 19
3',5'-bis-*O*-[tert-Butyl(dimethyl)silyl]-2'-deoxy-2'-(*R*)-fluoro-(N^3,N^4-$^{15}N_2$)-cytidine (13) 20
3',5'-bis-*O*-(tert-Butyl(dimethyl)silyl)-5-iodo-2'-deoxy-2'-(*R*)-fluoro-(N^3,N^4-$^{15}N_2$)-cytidine (14) 22
3',5'-bis-*O*-(tert-Butyl(dimethyl)silyl)-5-(methylcarboxy)-2'-deoxy-2'-(*R*)-fluoro-(N^3,N^4-$^{15}N_2$)-cytidine (15) 23
5-(Carboxy)-2'-deoxy-2'-(*R*)-fluoro-(N^3,N^4-$^{15}N_2$)-cytidine (3) 25
O-3',5'-bis(*tert*-Butyldimethylsilyl)-2'-deoxy-2'-(R)-fluoro-cytidine (16) 27
O-3',5'-bis(*tert*-Butyldimethylsilyl)-5-iodo-2'-deoxy-2'-(R)-fluoro-cytidine (17) 28
3'-5'-*O*-bis(*tert*-Butyldimethylsilyl)-5-methoxycarbonyl-2'-desoxy-2'-(R)-fluoro-cytidine (18) 30
3'-5'-*O*-bis(*tert*-Butyldimethylsilyl)-5-methoxycarbonyl-2'-desoxy-2'-(R)-fluoro-5,6-dihydrocytidine (19) 31
3'-5'-*O*-bis(*tert*-Butyldimethylsilyl)-2'-desoxy-2'-(R)-fluoro -5,6-dihydrocytidine (20) 33
Supplementary tables 34
Appendix 36
NMR Spectra 36

Supplementary figures

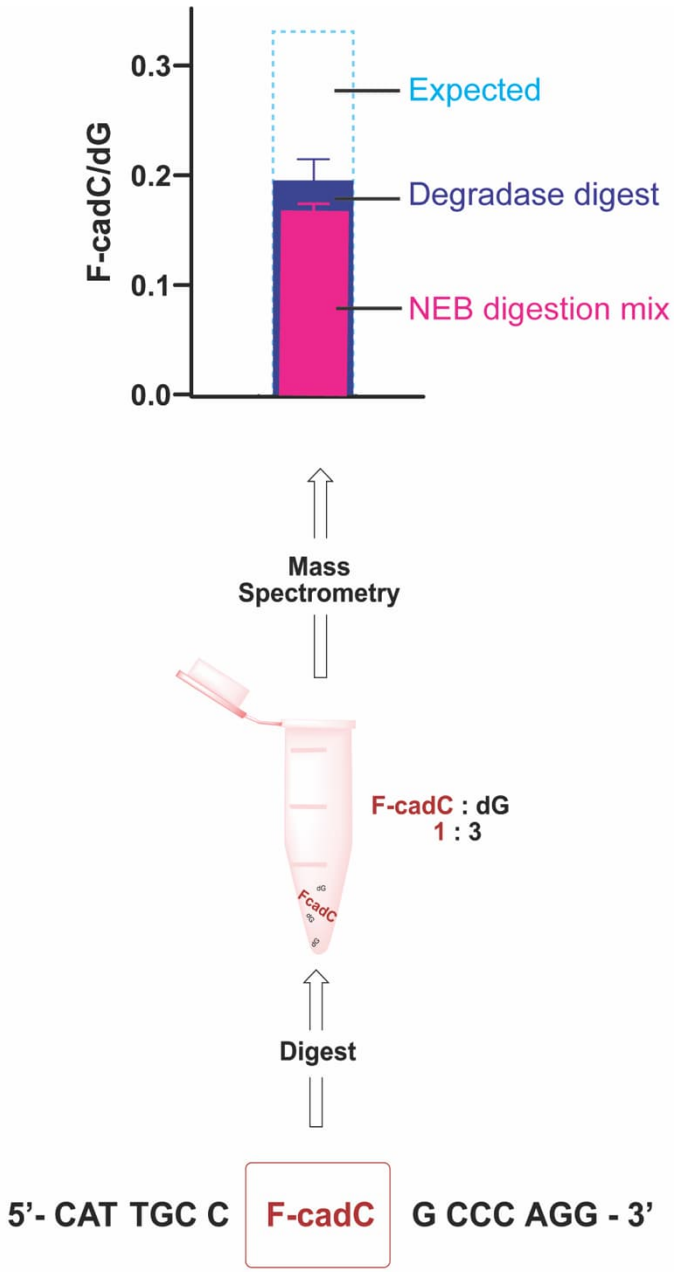


Figure SI-1: Digestion efficiency of the DNA oligonucleotides containing F-cadC using degradase mix (Zymo Research) when compared to NEB digestion mix (New England BioLabs). The degradase in our digestion test showed the highest efficiency and was used for all the experiments. The result is compliant with the literature[1].

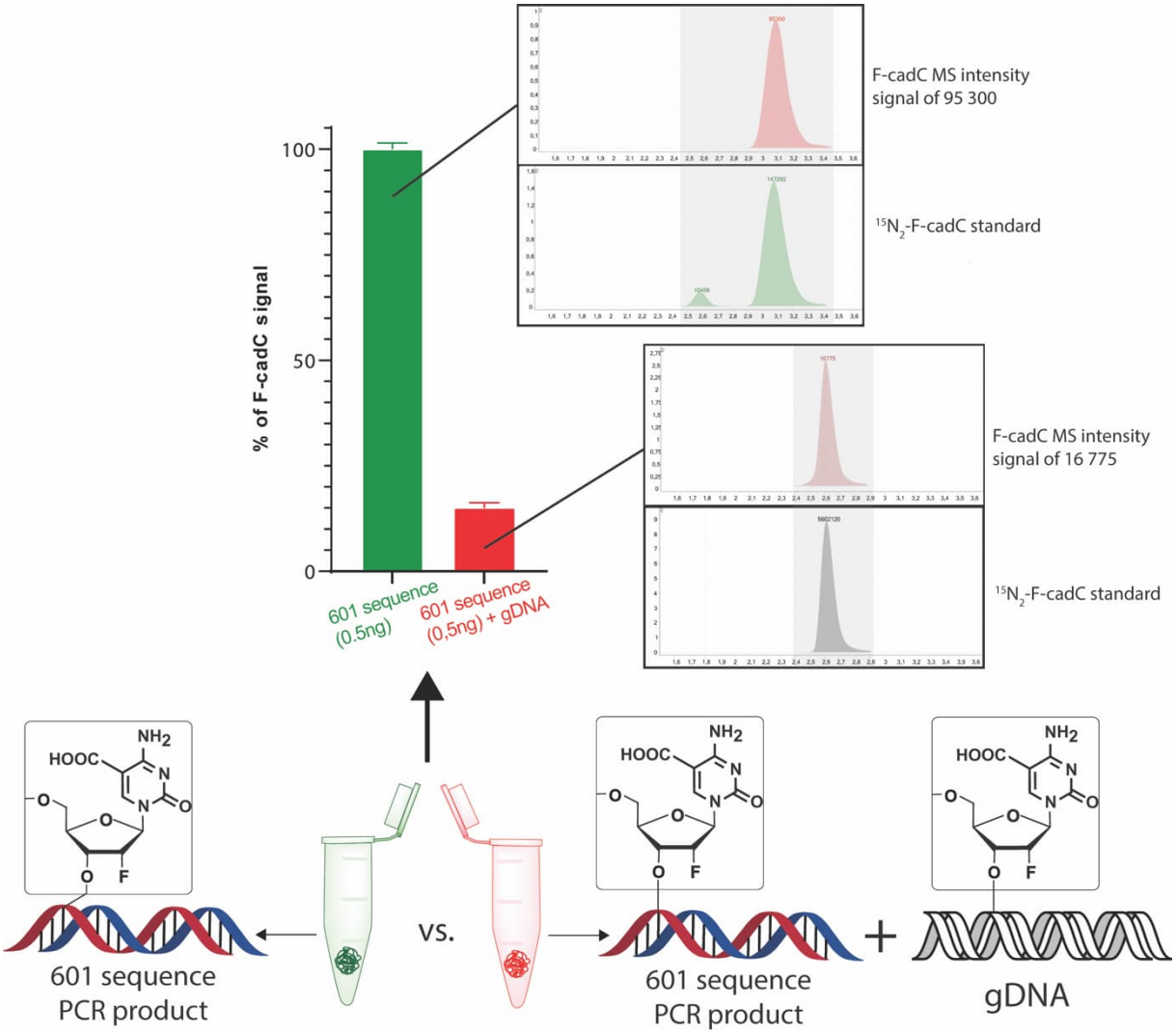


Figure SI-2: The graphical representation of the degree of signal suppression when F-cadC is present in the genomic DNA context. The green bar represents the signal intensity of the digested 0.5 ng of 601 Widom sequence[2] (containing 2'-F-cadC at 30 dC positions) on its own. The red bar shows 2'-F-cadC signal in the same sequence when digested together with 10μg of gDNA. The peaks are aligned with their isotopically labelled standards to confirm their identity. Any measurable quantity of the isotopically labelled standard when present within the sample also strongly supresses the signal, making it impossible to quantify the F-cadC amount using internal calibration and producing unreliable data when compared to external calibration standard curves.

601 sequence:
5´-ATCGATGTATATATCTGACACGTGCCTGGAGACTAGGGAGTAATCCCCTTGGCGGTTAAAA CGCGGGGGACAG-3´

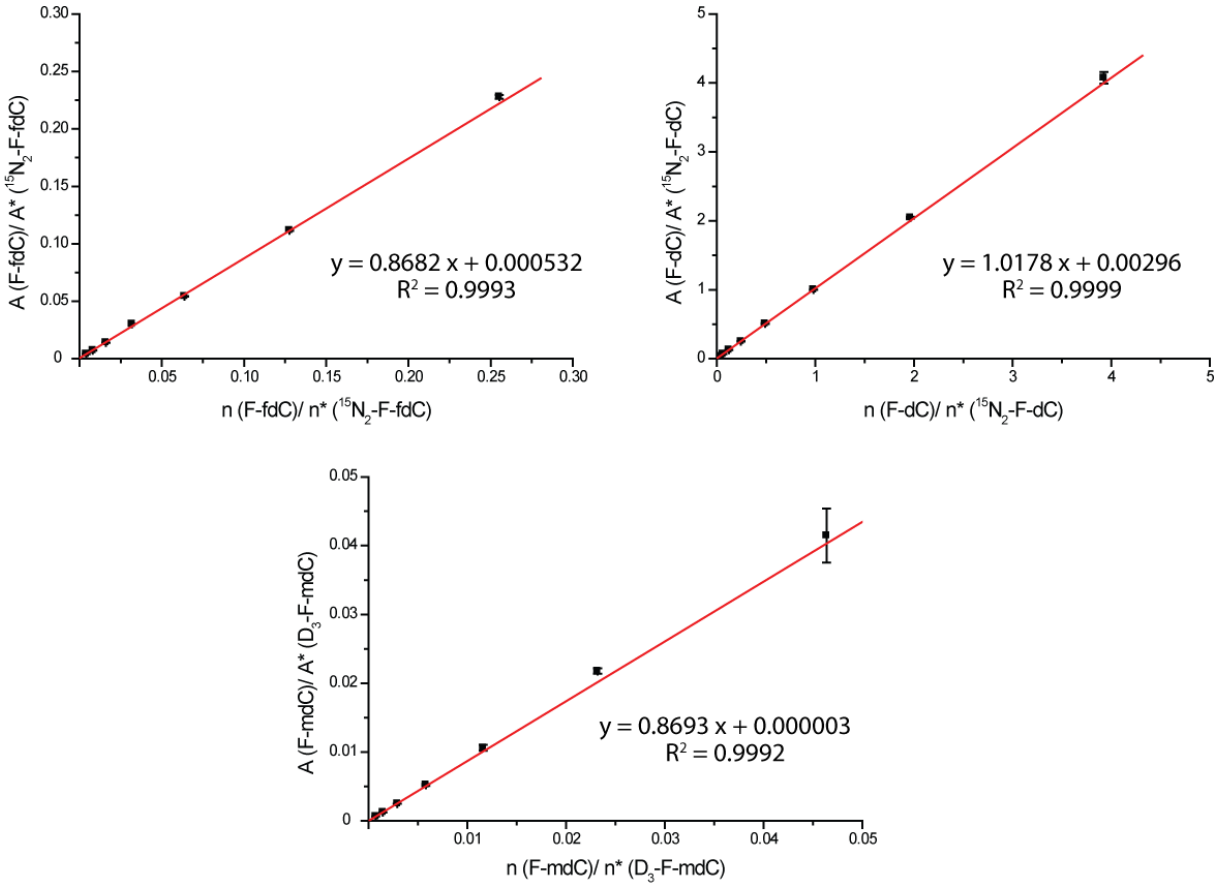


Figure SI-3: Internal calibration curves for the exact quantification of 2'-F-fdC (**5**), 2'-F-dC (**2**) and 2'-F-mdC (**7**) with the corresponding linear equation and coefficient of determination. [3]

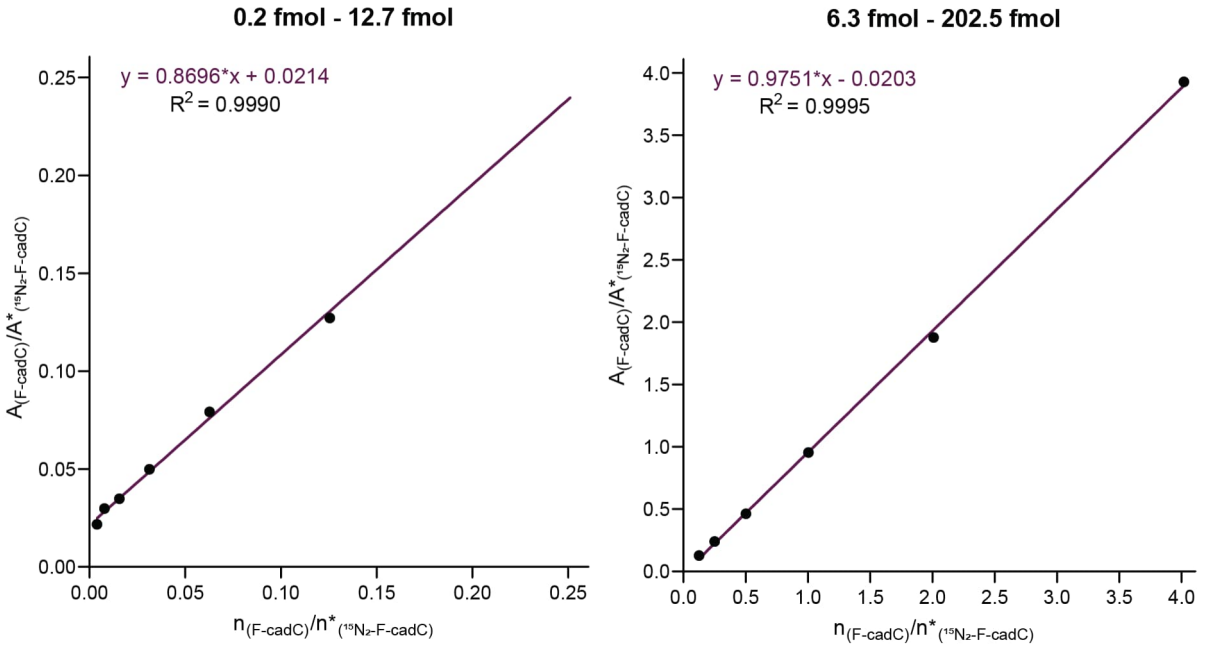


Figure SI-4: Internal calibration curves for the exact quantification of 2'-F-cadC (**1**), with the corresponding linear equation and coefficient of determination.

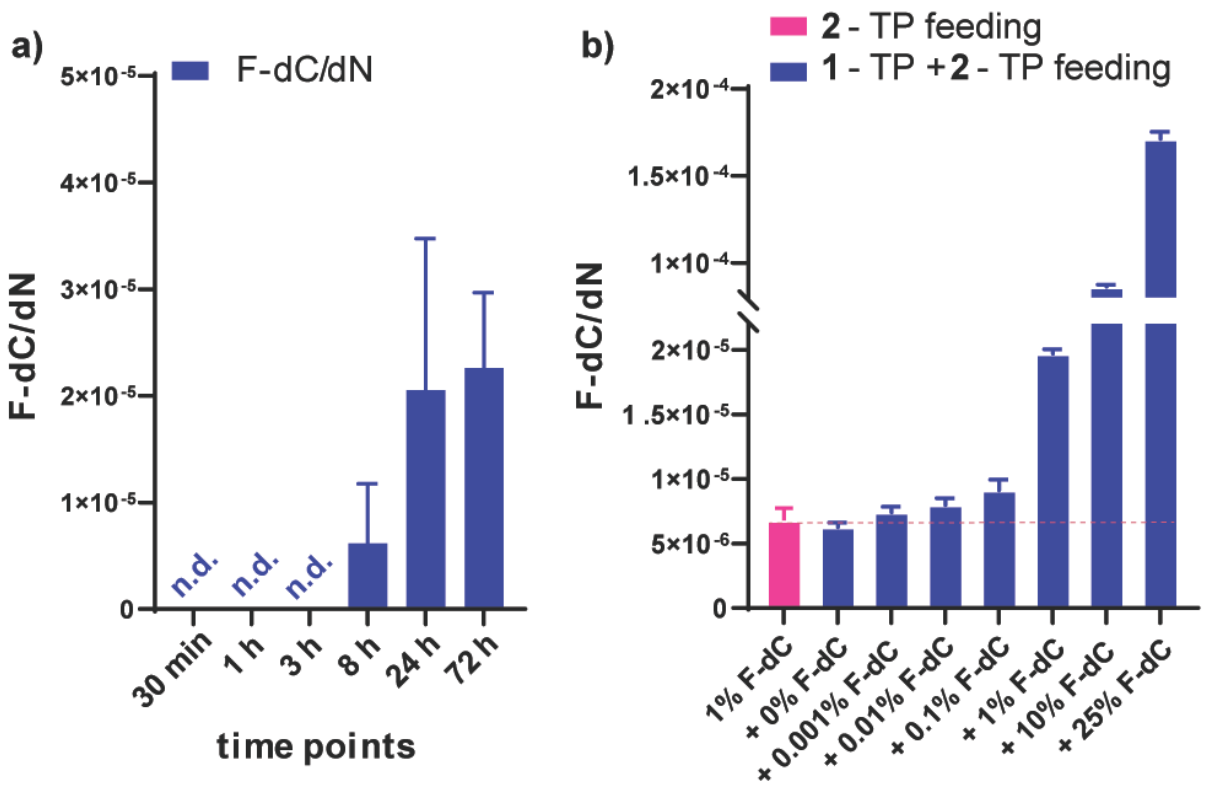


Figure SI-5: a) The pulse-chase study of the decarboxylation product 2'F-dC **2** formation upon 100μM of **1**-TP feeding to Neuro-2a cells for 30mins. Quantifiable amount of **2** are visible only after 8h from the start of the feeding. The error bars represent the propagated standard deviation of 3 biological replicates. b) The results of the control 24-hour experiment when different concentrations of "impurity" of **2**-TP were co-fed with 100μM of **1**-TP. The pink bar represents the cells fed only with 1% (1μM) **2**-TP. We can see that using 1μM of **2**-TP only ("1% F-dC") we get the incorporation rate marginally higher than in the feeding of 100μM of **1**-TP alone ("+0% F-dC"). In turn, when we combine both samples 1μM of **2**-TP and 100μM of **2**-TP ("+1% F-dC"), the intragenomic levels of **2** are about three times higher than in the 100μM **2**-TP ("+0% F-dC") feeding. The error bars show the standard deviation of a technical triplicate.

If decarboxylation was happening in the cytosol, not in the genome, we would have expected the turnover to happen to a higher extent than 1%. This result, together with the soluble pool investigation (Figure 3a in the main text) that shows no detectable F-dC in the cytosol hints at the observed decarboxylation as an intragenomic process.

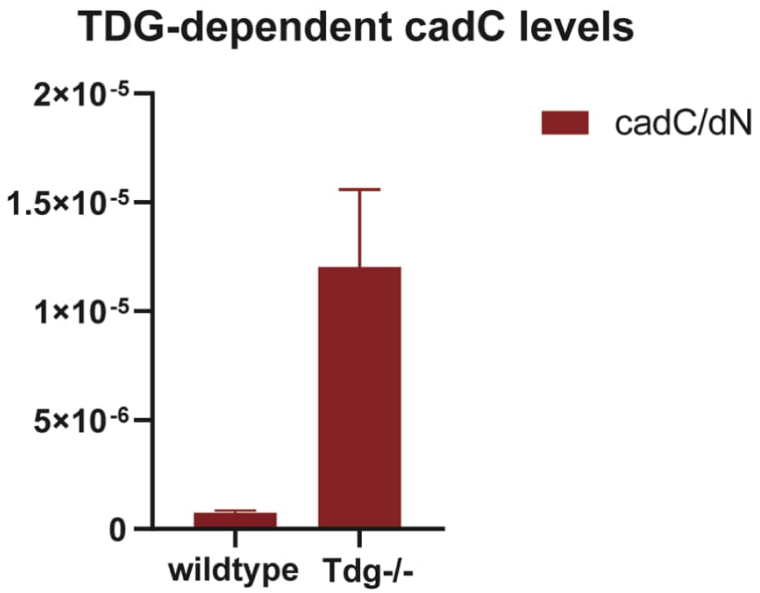


Figure SI-6: The natural cadC levels in the genome are usually very low in the presence of TDG. This was also the case when we measured cadC in the wt E14 mESCs. TDG knockout of the same cell line shows cadC levels two orders of magnitude higer than in the wildtype cells. This points out at the efficiency of the BER and assures us that in order to investigate the direct decarboxylation, we need a model system that excludes Tdg function. Using fluorine-modified nucleosides allows us to achieve this goal. From our previous studies, we know already that the levels of intragenomic 2'F-dC (**2**) formed upon 2'F-fdC (**5**) feeding shows the same result for both: wt and Tdg-/- cells. This confirms the lack of 2'F-fdC recognition by Tdg. The error bars represent the propagated error of the standard deviation.

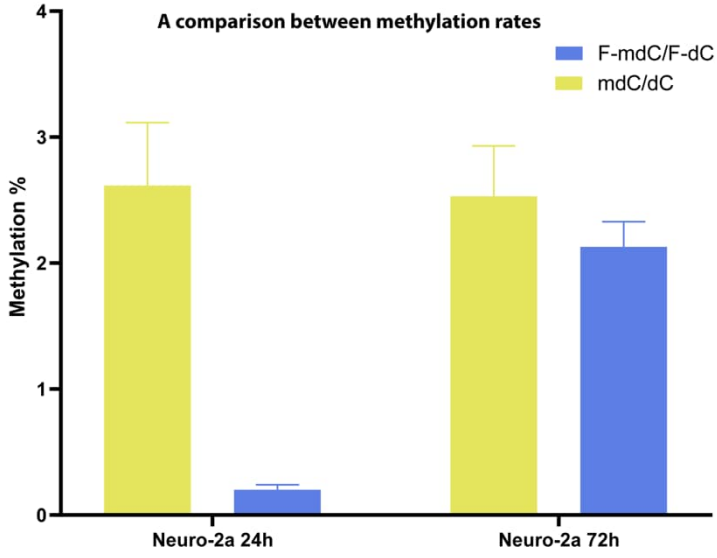


Figure SI-7: The methylation levels of natural dC and F-dC in the Neuro-2a cell line upon 24h and 72h feeding of 100μM F-cadCTP. Since the natural methylation levels show values ~2.5%, confirmed also in the unfed control, the F-mdC/F-dC reach about 0.2% after 24h and 2.1% after 72h. This data confirms that shorter feedings do not allow us to see full methylation, however, after three days we reach the values close to the natural methylation, thus decarboxylation happens in the CpG and non-CpG context. The error bars represent standard deviation of three technical replicates.

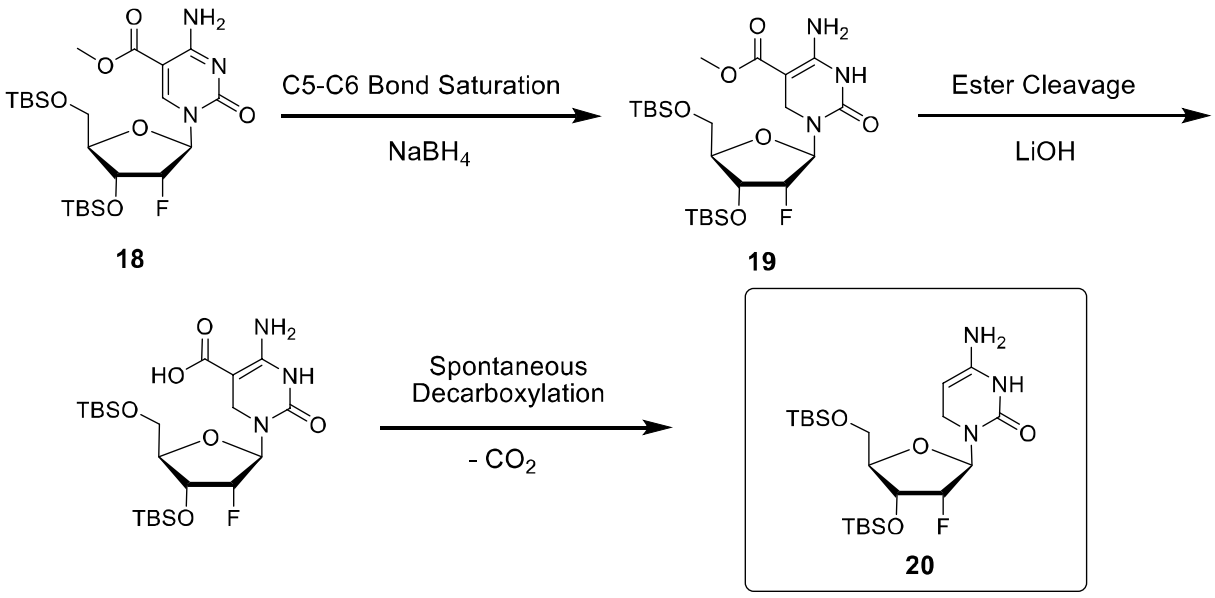


Figure SI-8: Spontaneous Decarboxylation of TBS-protected, 5-6 saturated F-cadC (Synthesis below). F-dC is first TBS protected with TBSCl and imidazole, before being iodinated with I_2 and oxidiser CAN. Afterwards, CO-insertion yields TBS protected 5-methoxycarbonyl-FdC, which is then saturated at the C5-C6 bond with $NaBH_4$. Ester cleavage with LiOH then leads to spontaneous decarboxylation towards TBS-protected 5,6-saturated FdC **20**.

Methods and Materials

Chemical Synthesis

Unless noted otherwise, all chemical transformations were performed using flame- or oven dried glassware under an atmosphere of argon. Chemicals were purchased from *Sigma-Aldrich*, *TCI*, *Fluka*, *ABCR*, *Carbosynth* or *Acros Organics* and used without further purification. Technical grade solvents were distilled prior to extraction or chromatography. Reactions and column chromatography fractions were monitored by thin-layer chromatography (TLC) on silica gel F_{254}TLC plates from Merck KGaA. Visualization of the developed TLC plates was achieved through UV-absorption or through staining with Hanessian's stain or $KMnO_4$ stain. Flash column chromatography was performed on Geduran®Si60 (40-63 µm) silica gel from Merck KGaA applying slight nitrogen pressure. NMR spectra were recorded on *Bruker Avance III HD 400 (*400 MHz), *Varian NMR-System 600* (600 MHz) and *Bruker Avance III HD* with *Cryo-Kopf 800* (800 MHz) spectrometers. ^{1}H-NMR shifts were calibrated to the residual solvent resonances: $CDCl_3$ (7.26 ppm), DMSO-d_6 (2.50ppm) and D_2O (4.79 ppm). ^{13}C-NMR shifts were calibrated to the residual solvent: $CDCl_3$ (77.16 ppm), DMSO-d_6 (39.52 ppm). All NMR spectra were analysed using the program MestReNova 10.0.1 from Mestrelab NMR data are reported as follows: chemical shift (multiplicity, coupling constants where applicable, number of hydrogens, assignment). Splitting is reported with the following symbols: s = singlet, d = doublet, t = triplet, dd = doublet of doublets, dt = doublet of triplets, ddd = doublet of doublets of doublets, m = multiplet. Whenever a signal is not clearly distinguishable from related nuclides, asterisks (e.g. *, **) are attached to indicate interchangeability. Research S. L. Low resolution mass spectra were measured on a LTQ FT-ICR by Thermo Finnigan GmbH. High resolution mass spectra were measured by the analytical section of the Department of Chemistry of Ludwigs-Maximilians-Universität München on aMAT 90 (ESI) from Thermo Finnigan GmbH. IR spectra were recorded on a PerkinElmer Spectrum BX II FT-IR system. Substances were applied as a film or directly as solids on the ATR unit. Analytical RP-HPLC was performed on an analytical HPLC Waters Alliance (2695 Separation Module, 2996 Photodiode Array Detector) equipped with the column Nucleosil120-2 C18 from Macherey Nagel applying an eluent flow of 0.5 mL/min. Preparative RP-HPLC was performed on HPLC Waters Breeze (2487 Dual λ Array Detector, 1525 Binary HPLC Pump) equipped with the column VP 250/32 C18 from Macherey Nagel. A flowrate of 5mL/min was applied. HPLC grade MeCN was purchased from VWR.

Cell culture

DMEM high glucose (Sigma Aldrich) containing 10% heat-inactivated FBS (Gibco), 100 U/mL Penicillin-Streptomycin (Gibco), 2 mM L-glutamine were used to culture rat neuroblastoma Neuro-2a cells.

DMEM high glucose and Ham's Nutrient Mixture F12 (Sigma Aldrich) mixture 50/50 containing 100 U/mL Penicillin-Streptomycin (Gibco), containing 10% heat-inactivated FBS (Gibco), 2 mM L-glutamine were used to culture CHO-K1 cells.

mES cells (J1, E14, R1) were cultured in three different media – slitting, maintenance (a2i) and priming (C/R) media. The maintenance medium keeps stem cells in naïve pre-implantation state. It is consisted of DMEM high glucose (Sigma Aldrich), 10% heat-inactivated FBS (Gibco), 0.1mM β-mercaptoethanol (Sigma Aldrich), 2mM L-Alanyl-L-Glutamine 200mM (Gibco), 1x Nonessential amino acids 100x (Sigma Aldrich) and 100 U/mL Penicillin-Streptomycin (Gibco). The culture medium consists of the splitting medium supplemented with LIF 1000U/mL (ORF Genetics), 3.0 μM CGP 77675 (Sigma Aldrich) and 3.0 μM GSK3 inhibitor CHIR-99021 (Sigma Aldrich). The priming media consisted of the splitting media supplemented with LIF 1000U/mL (ORF Genetics), 2.5 μM tankyrase inhibitor IWR-1 endo (Sigma Aldrich) and 3.0 μM CHIR-99021 (Sigma Aldrich). The cells were primed for 3 days via plating in C/R medium and fed with the **1**-TP during the last 24 hours of priming.

Metabolic-labelling experiments

Triphosphate feedings were conducted by plating cells in their dedicated medium for 24h in 6 cm petri dishes to obtain a confluency of 70-80% prior to adding the compound. The NTP transporter was purchased from Merck and resuspended to a stock solution of 5mM with autoclaved ddH_2O. The feeding was performed in a ratio of 1:1 transporter to triphosphate concentration in tricine buffer (prepared according to manufacturer's protocol) in the total volume of 450 μL. The medium was first removed from the dish, the cells were washed once with PBS (Sigma Aldrich) and the transporter + PPP mixture was applied for 30 minutes for somatic cells and 15mins for the stem cells at 37°C, 5% CO_2. Afterwards the cells were again washed with PBS and fresh medium was applied.

The study of the

Cell Harvesting and Isolation of genomic DNA

All samples were washed with PBS (Sigma Aldrich) once before harvesting and lysed directly in the plates by adding RLT buffer (Qiagen) supplemented with 400µM 2,6-di-tert-butyl-4-methylphenol (BHT) and desferoxamine mesylate (DM) both from Sigma Aldrich. Next, DNA was sheered in MM400 bead mill (Retsch) at 30 Hz for 1 min in 2mL microcentrifuge tubes (Eppendorf) with 5mm diameter stainless steel beads (one per tube) and centrifuged at 21000rcf for 5 minutes. Genomic DNA (gDNA) was extracted using Zymo Quick gDNA mini-prep® kit according to the protocol with an addition of RNAse A treatment (Qiagen) at 0.2mg/mL in Genomic Lysis Buffer – 400µL of the solution transferred directly on the column and incubated for 15 minutes. All samples were eluted in DNAse-free ddH_2O with 20µM BHT, the concentration of gDNA was measured on Nanodrop.

DNA digestion

While investigating the decarboxylation *in vivo* using 2´-F-cadC-TP (**1**-TP) as a metabolic label, the following conditions were used for the gDNA digestion procedure: 1 – 10 µg of genomic DNA or oligonucleotide in 35 µL H_2O were digested as follows: 2 µL of Degradase plus (*Zymo research*), 5 µL of Degradase plus reaction buffer (*Zymo research*), 0.5 µL of Benzonase nuclease (*Merck, Serratia marcescens*) and a specific amount of isotopically labelled internal standards were added. Before adding the digestion mixture, the gDNA was heated to 95°C for 5 minutes and placed directly on ice. The mixture was incubated for 4 h at 37 °C, diluted with 450 µL of water and then extracted with chloroform (500 µL) twice. The samples were lyophilised and resuspended in 50 µL ddH_2O. Prior to UHPLC-MS/MS analysis, samples were filtered by using an AcroPrep Advance 9 filter plate 0.2 µm Supor (*Pall Life Science*).
For digestion of DNA samples using the Nucleoside Digestion Mix from New England BioLabs, a solution of 1 - 10 µg genomic DNA or oligonucleotide in 35 µL of water was incubated with 5 µL of the Nucleoside Digestion Mix Reaction Buffer (10x), 2.5 µL of the Nucleoside Digestion Mix and a specific amount of isotopically labelled internal standards for 2 h at 37 °C. The sample was filtered using an AcroPrep Advance 96 filter plate 0.2 µm Supor from Pall Life Sciences and subsequently analyzed by UHPLC-MS/MS.

LC/MS-MS analysis of DNA samples[4]

Quantitative UHPLC-MS/MS analysis of digested DNA samples was performed using an Agilent 1290 UHPLC system equipped with a UV detector and an Agilent 6490 triple quadrupole mass spectrometer. Nucleosides of interest were quantified using the stable isotope dilution technique.[3] The source-dependent parameters were as follows: gas temperature 80 °C, gas flow 15 L/min (N_2), nebulizer 30 psi, sheath gas heater 275 °C, sheath gas flow 11 L/min (N2), capillary voltage 2.500 V in the positive ion mode, capillary voltage −2.250 V in the negative ion mode and nozzle voltage 500 V. The fragmentor voltage was 380 V/ 250 V. Delta EMV was set to 500 (positive mode) and 800 (negative mode). Compound-dependent parameters are summarized in Supplementary Tables 1-3. Chromatography was performed by a Poroshell 120 SB-C8 column (Agilent, 2.7 µm, 2.1 mm × 150 mm) at 35 °C using a gradient of water and MeCN, each containing 0.0085% (v/v) formic acid, at a flow rate of 0.35 mL/min: 0 →4 min; 0 →3.5% (v/v) MeCN; 4 →7.9 min; 3.5 →5% MeCN; 7.9 →8.2 min; 5 →80% MeCN; 8.2 →11.5 min; 80% MeCN; 11.5 →12 min; 80 →0% MeCN; 12 →14 min; 0% MeCN. The autosampler was cooled to 4 °C. The injection volume was amounted to 39 µL.

Quantification of nucleosides in DNA samples

Nucleosides were quantified using internal calibration curves and the stable isotope dilution technique as described in the literature. [4] Used internal calibration curves for F-nucleosides are shown in Supplementary Figures 3 and 4.

Soluble nucleoside pool extraction and purification

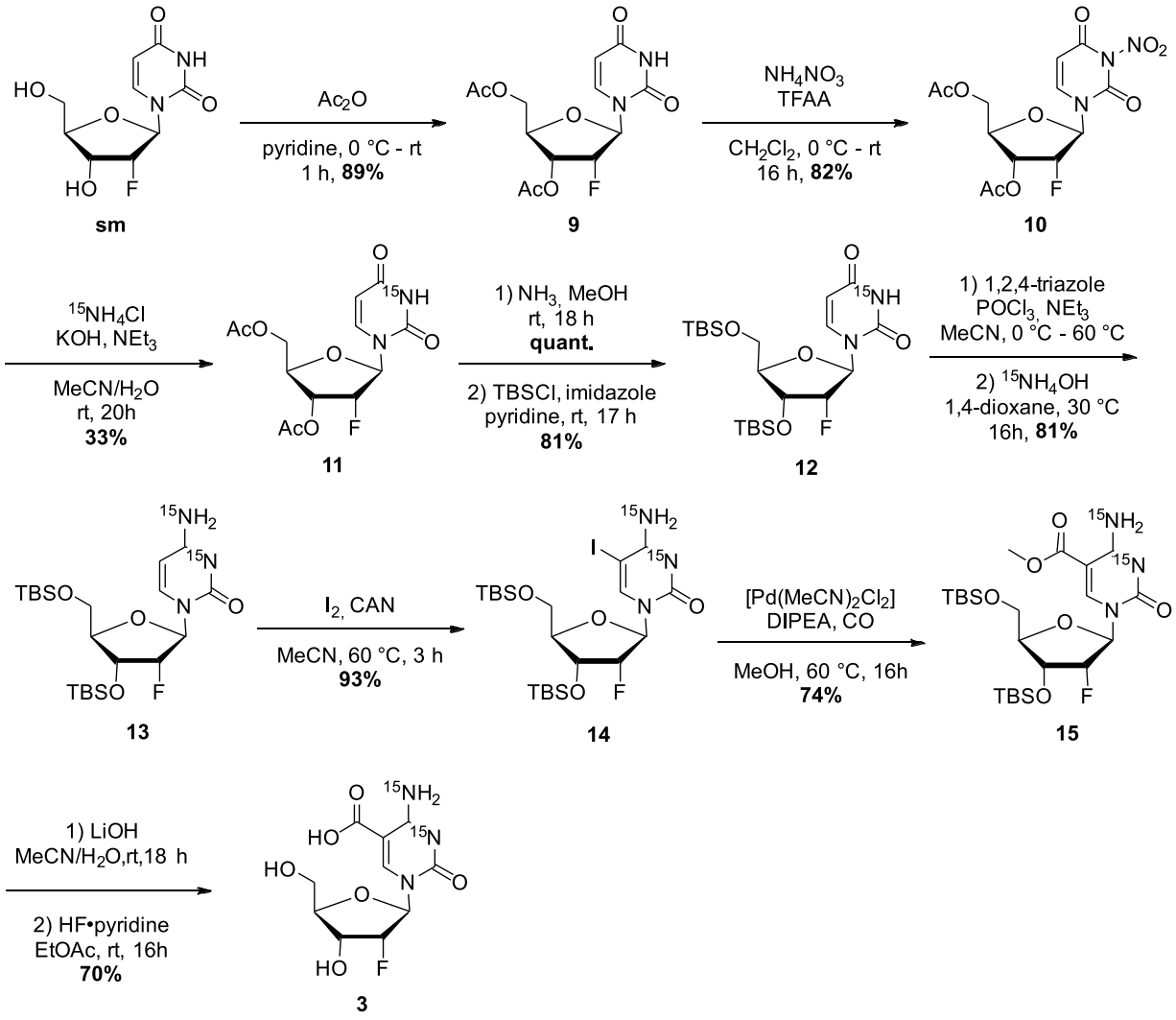
The soluble pool extraction protocol started with a cell harvest using typical cell passaging trypsinisation technique using TrypLE reagent, Gibco. The cells were counted with trypan blue via Countess II FL Automated Cell Counter, Life Technologies (ThermoFisher Scienitfic) and pelleted via centrifugation at 260g for 3min. The pellet was resuspended with the ice-cold mixture of acetonitrile and water (1:1 ratio), 1mL per $1x10^6$cells and gently vortexed. The cell suspension was incubated on ice for 15min and centrifuged at 10 000g for 10min at 4°C. The supernatant containing the cytosolic fraction was collected and lyophilised, while the cell pellet was treated with the lysis buffer for subsequent gDNA isolation.

Samples were dissolved in H_2O (100 µL H_2O for soluble pool from 400 000 cells) and treated with Antarctic Phosphatase (New England BioLabs) for 30 min. Supel-Select HBL 60 mg columns were equilibrated with 1.5 mL MeOH and 3 mL H_2O (pH = 4). Samples were loaded

the surface of the column, washed with 4 mL H_2O (pH = 4). Cartridges were dried under high vacuum for 15 min. Sample was eluted with 1.5 mL MeOH/MeCN = 1:1. Eluent was lyophilized, dissolved in H_2O and analysed by UHPLC-MS/MS as described for genomic DNA samples.[3]

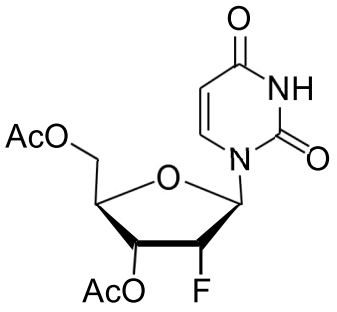
Synthetic procedures

The nucleoside triphosphates were synthesised according to literature. [5-6]



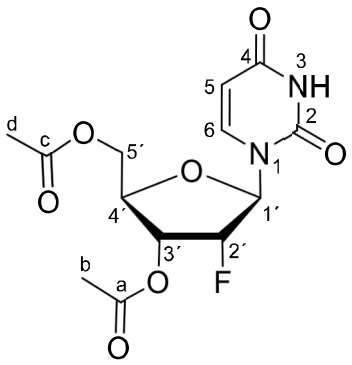
Scheme SI-1: Synthesis pathway towards $^{15}N_2$-2'-F-cadC (**3**).

3',5'-di-*O*-Acetyl-2'-deoxy-2'-(*R*)-fluorouridine (9)



9

2‘-Deoxy-2'-fluorouridine (2.40 g, 9.75 mmol, 1.0 equiv.) was dissolved in pyridine (100 mL), cooled to 0 °C and acetic anhydride (19.2 mL, 68.3 mmol, 7.0 equiv.) was added. The reaction mixture was stirred for 1 h at 0 °C, warmed to room temperature and stirred for additional 16 h. Afterwards, pyridine was removed under reduced pressure, the resulting yellow oil was coevaporated with toluene (3 x 200 mL) and taken up in an aqueous, saturated $NaHCO_3$-solution (100 mL). The solution was extracted with CH_2Cl_2 (3 x 150 mL), the combined organic layers were washed with an aqueous, saturated NaCl solution (300 mL) and dried over Na_2SO_4. The mixture was concentrated *in vacuo*. The resulting crude was purified by flash chromatography (SiO_2, CH_2Cl_2 → CH_2Cl_2/MeOH = 99:1) to yield compound **9** as a colourless solid (2.85 g, 8.63 mmol, 89%).



R_f = 0.72 (SiO_2, CH_2Cl_2/MeOH = 20:1).

^{1}H-NMR (600 MHz, $CDCl_3$): δ/ppm = 8.24 (s, 1H, C3-H), 7.39 (d, ^3J = 8.2 Hz, 1H, C6-H), 5.88 (dd, $^3J_{H\text{-}F}$ = 19.4 Hz, ^3J = 1.5 Hz 1H, C1'-H), 5.76 (dd, ^3J = 8,2 Hz, ^3J = 2.2 Hz, 1H, C5-H), 5.37 (ddd, $^2J_{H\text{-}F}$ = 52.2 Hz, ^3J = 5.0 Hz, ^3J = 1.5 Hz, 1H, C2'-H), 5.15 (ddd, $^3J_{H\text{-}F}$ = 17.9 Hz, ^3J = 8.2 Hz, ^3J = 5.0 Hz, 1H, C3'-H), 4.45 (dd, ^2J = 12.4 Hz, ^3J = 2.5 Hz, 1H, C5'-H), 4.43–4.40 (m, 1H, C4'-H), 4.31 (dd, ^2J = 12.4 Hz, ^3J = 4.5 Hz, 1H, C5'-H), 2.17 (s, 3H, C_d-H/C_b-H)*, 2.11 (s, 3H, C_d-H/C_b-H)*.

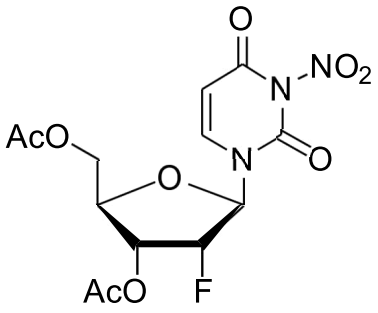
^{13}C-NMR (101 MHz, $CDCl_3$): δ/ppm = 170.6 (C_c), 170.2 (C_a), 162.6 (C4), 149.8 (C2), 140.9 (C6), 103.4 (C5), 91.7 (d, $^2J_{C\text{-}F}$ = 35.6 Hz, C1'), 91.3 (d, $^1J_{C\text{-}F}$ = 191.9 Hz, C2'), 79.1 (C4'), 70.0 (d, $^2J_{C\text{-}F}$ = 15.6 Hz, C3'), 62.6 (C5'), 21.2 (C_b/C_d)*, 20.8 (C_b/C_d)*.

^{19}F-NMR (126 MHz, $CDCl_3$): δ/ppm = -199.1 (dt, $^2J_{H\text{-}F}$ = 52.3 Hz, $^3J_{H\text{-}F}$ = 18.7 Hz, C2'-*F*).

HRMS (ESI): calc. for $C_{13}H_{16}FN_2O_7^+$ $[M+H]^+$: 331.09361, found: 331.09363.

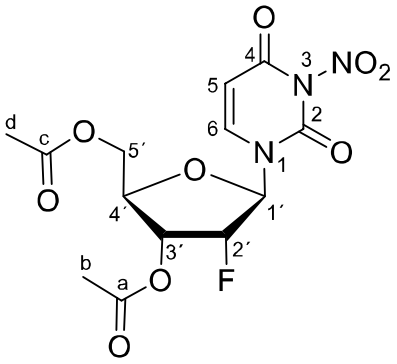
The data are consistent with literature.[3]

3',5'-di-*O*-Acetyl-2'-deoxy-2'-(*R*)-fluoro-3-nitrouridine (10)



10

NH_4NO_3 (1.84 g, 23.0 mmol, 2.0 equiv.) was ground with a mortar, suspended in CH_2Cl_2 (65 mL) and cooled to 0 °C for 20 min before trifluoroacetic anhydride (6.49 mL, 46.0 mmol, 4.0 equiv.) was added dropwise. The reaction mixture was stirred for 3 h at room temperature until the remaining NH_4NO_3 was dissolved. After the resulting yellow solution was cooled to 0 °C, compound **9** (3.80 g, 11.5 mmol, 1.0 equiv.) was added in portions. The reaction mixture was stirred for 3 h at 0 °C, warmed to room temperature and stirred for additional 16 h. After complete conversion of the nucleoside, the mixture was neutralised by addition of PBS-buffer (130 mL) and the aqueous solution was extracted with CH_2Cl_2 (3 x 100 mL). The combined organic layers were dried over Na_2SO_4 and solvents were removed *in vacuo*. The resulting yellowish oil was purified by flash chromatography (SiO_2, CH_2Cl_2/MeOH = 50:1) to yield nucleoside **10** as a colourless solid (3.52 g, 9.38 mmol, 82%).



$\mathbf{R_f}$ = 0.44 (SiO_2, CH_2Cl /MeOH = 9:1).

^{1}H-NMR (400 MHz, $CDCl_3$): δ/ppm = 7.46 (d,^3J = 8.4 Hz, 1H, C6-H), 5.90 (d,^3J = 8.4 Hz, 1H,C5-H), 5.81 (dd, $^3J_{H\text{-}F}$ = 18.7 Hz, ^3J = 1.6 Hz, 1H, C1'-H), 5.42 (ddd, $^2J_{H\text{-}F}$ = 51.8 Hz, ^3J = 4.9 Hz, ^3J = 1.6 Hz, 1H, C2'-H), 5.14 (ddd, $^3J_{H\text{-}F}$ = 18.0 Hz, ^3J = 8.4 Hz, ^3J = 4.9 Hz, 1H, C3'-H), 4.49 – 4.42 (m, 2H, C5'-H, C4'-H), 4.36 – 4.29 (m, 1H, C5'-H), 2.17 (s, 3H, C_b-H/C_d-H)*, 2.11 (s, 3H, C_b-H/C_d-H)*.

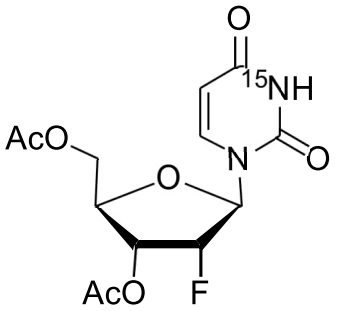
^{13}C-NMR (101 MHz, $CDCl_3$): δ/ppm = 170.2 (C_c), 169.8 (C_a), 154.8 (C4), 145.1 (C2), 140.1 (C6), 102.1 (C5), 92.1 (d, $^2J_{C\text{-}F}$ = 36.2 Hz, C1'), 90.8 (d, $^1J_{C\text{-}F}$ = 193.2 Hz, C2'), 79.3 (C4'), 69.3 (d, $^2J_{C\text{-}F}$ = 15.6 Hz, C3'), 61.8 (C5'), 20.8 (C_b/C_d)*, 20.5 (C_b/C_d)*.

^{19}F-NMR (376 MHz, $CDCl_3$): δ/ppm = 199.3 (dt, $^2J_{H\text{-}F}$ = 51.8 Hz, $^3J_{H\text{-}F}$ = 18.4 Hz, C2'-*F*).

HRMS (ESI): calc. for $C_{13}H_{18}FN_4O_9^+$ $[M+NH_4]^+$: 393.10523, found: 393.10527.

The data are consistent with literature. [3]

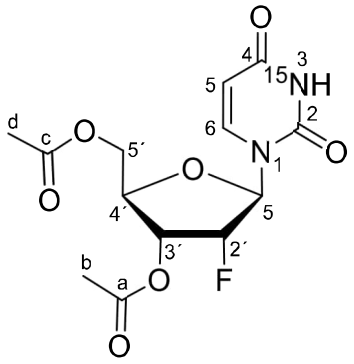
3',5'-di-*O*-Acetyl-2'-deoxy-2'-(*R*)-fluoro-(*N*^{3}-^{15}N)-uridine (11)



11

$^{15}NH_4Cl$ (0.261 g, 4.79 mmol, 1.3 equiv.) was dissolved in milli-Q-H_2O (10 mL), KOH (0.269 g, 4.79 mmol, 1.3 equiv.) and NEt_3 (0.89 mL, 6.39 mmol, 1.8 equiv.) were added. A solution of compound **10** (1.36 g, 3.62 mmol, 1.0 equiv.) in MeCN (10 mL) was added to the reaction

mixture and stirred for 20 h at room temperature. After complete conversion, MeCN was removed *in vacuo*. The resulting solution was neutralised with aqueous, saturated NH_4Cl (15 mL) and extracted with CH_2Cl_2 (3 x 100 mL). The combined organic layers were dried over Na_2SO_4 and concentrated under reduced pressure. The resulting colourless crude was purified by flash chromatography (SiO_2, CH_2Cl_2 → CH_2Cl_2/MeOH = 97:3). Compound **11** was obtained as a colourless solid (0.402 g, 1.21 mmol, 33%).



R_f = 0.38 (SiO_2, CH_2Cl_2/MeOH = 99:1).

^{1}H-NMR (400 MHz, $CDCl_3$): δ/ppm = 8.29 (dd, $^2J_{H-^{15}N}$ = 91.2 Hz, ^{4}J = 2.2 Hz, 1H, N3-H), 7.39 (d,^3J = 8.2 Hz, 1H, C6-H), 5.83 – 5.74 (m, 1H, C1'-H), 5.77 (d, ^3J = 8.2 Hz, 1H, C5-H), 5.37 (ddd, $^2J_{H-F}$ = 52.2 Hz, ^3J = 4.8 Hz, ^3J = 1.6 Hz, 1H, C2'-H), 5.15 (ddd, $^3J_{H-F}$ = 17.9 Hz, ^3J = 5.3 Hz, ^3J = 2.8 Hz, 1H, C3'-H), 4.47 – 4.43 (m, 1H, C5'-H), 4.41 (dd, ^3J = 8.0 Hz, ^3J = 2.6 Hz, 1H, C4'-H), 4.31 (dd, ^2J = 12.1 Hz, ^3J = 4.2 Hz, 1H, C5'-H), 2.16 (s, 3H, C_b-H/C_d-H)*, 2.12 (s, 3H, C_b-H/C_d-H)*.

^{13}C-NMR (101 MHz, $CDCl_3$): δ/ppm = 170.3 (C_c), 169.9 (C_a), 162.3 (sd, $^1J_{C-N}$ = 9.2 Hz, C4), 149.5 (sd, $^1J_{C-N}$ = 18.3 Hz, C2), 140.6 (C6), 103.1 (sd, $^1J_{C-N}$ = 7.2 Hz, C5), 91.8 (d, $^2J_{C-F}$ = 38.3 Hz, C1'), 90.7 (d, $^1J_{C-F}$ = 118.0 Hz, C2'), 78.8 (C4'), 69.8 (d, $^2J_{C-F}$ = 15.6 Hz, C3'), 62.3 (C-5'), 20.9 (C_b/C_d)*, 20.5 (C_b/C_d)*.

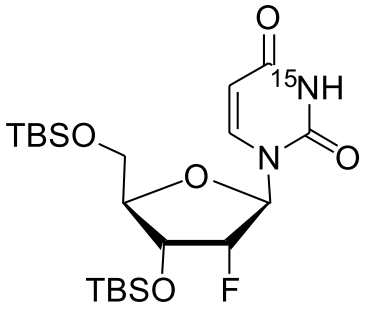
^{19}F-NMR (376 MHz, $CDCl_3$): δ/ppm = -199.1 (dt, $^2J_{H-F}$ = 52.3 Hz, $^3J_{H-F}$ = 18.7 Hz, C2'-*F*).

^{15}N-NMR (41 MHz, $CDCl_3$): *δ* (ppm) = -224.6.

HRMS (ESI): calc. for $C_{13}H_{14}FN^{15}NO_7^-$ [M-H]$^-$: 330.07609, found: 330.07654.

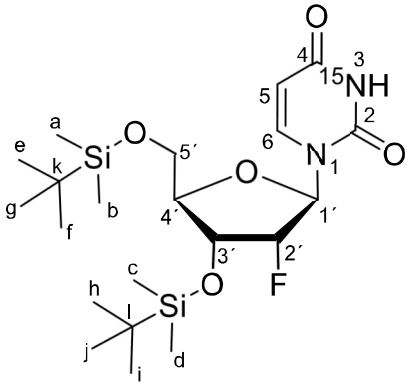
The data are consistent with literature. [3]

3',5'-bis-*O*-(tert-Butyl(dimethyl)silyl)-2'-deoxy-2'-(*R*)-fluoro-(*N*^{3}-^{15}N)-uridine (12)



12

Compound **11** (0.875 g, 2.64 mmol, 1.0 equiv.) was dissolved in ammonia (7 N in methanol, 5.5 mL) and stirred for 18 h at room temperature. Solvents were removed *in vacuo* and the residue dried under high vacuum after which the compound was processed without further purfiication. The nucleoside was dissolved in pyridine (6 mL) and imidazole (0.809 g, 11.9 mmol, 4.5 equiv.) and TBSCl (1.19 g, 7.92 mmol, 3.0 equiv.) were added. The mixture was stirred for 17 h at room temperature. Pyridine was removed under reduced pressure; the resulting residue was taken up in an aqueous $NaHCO_3$ solution (100 mL) and extracted with CH_2Cl_2 (3 x 150 mL). The combined organic layers were washed with an aqueous, saturated NaCl solution (100 mL) and dried over Na_2SO_4. Solvents were removed *in vacuo* and the resulting yellow oill was coevaporated with toluene (3 x 50 mL). The TBS-protected nucleoside **12** was yielded as a colourless, waxy solid (1.02 g, 2.14 mmol, 81%).



R_f = 0.49 (SiO_2, CH_2Cl_2/MeOH = 99:1).

^{1}H-NMR (600 MHz, $CDCl_3$): δ/ppm = 8.10 (d, ^3J = 7.6 Hz, 1H, C6-H), 6.04 (d,$^3J_{H\text{-}F}$ = 15.5 Hz, 1H, C1'-H), 5.96 (d, ^3J = 7.6 Hz, 1H, C5-H), 4.79 (dd, $^2J_{H\text{-}F}$ = 52.2 Hz, ^3J = 4.0 Hz, 1H, C2'-H), 4.23 (ddd, $^3J_{H\text{-}F}$ = 21.6 Hz, ^3J = 8.0 Hz, ^3J = 4.0 Hz, 1H, C3'-H), 4.13–4.07 (m, 2H, C4'-H, C5'-H), 3.79 (d, ^2J = 10.8 Hz, 1H, C5'-H), 0.94 (s, 9H, $C_{e\text{-}g}$-H)*, 0.89 (s, 9H, $C_{h\text{-}j}$-H)*, 0.12 (s, 6H, C_a-H, C_b-H), 0.12 (s, 3H, C_c-H)*, 0.10 (s, 3H, C_d-H)*.

^{13}C-NMR (201 MHz, $CDCl_3$): δ/ppm = 162.7 (C4), 149.9 (C2), 140.0 (C6), 102.6 (C5), 93.3 (d, $^1J_{C\text{-}F}$ = 196.6 Hz, C2'), 88.0 (d, $^2J_{C\text{-}F}$ = 33.8 Hz, C1'), 84.1 (C4'), 68.9 (d, $^2J_{C\text{-}F}$ = 16.6 Hz, C3'), 61.0 (C5'), 26.2 ($C_{e\text{-}g}$)*, 25.9 ($C_{h\text{-}j}$)*,18.7 (C_l)**, 18.4 (C_k)**, -4.3 (C_a)***, -4.7 (C_b)***, -5.1 (C_c)***, -5.2 (C_d)***.

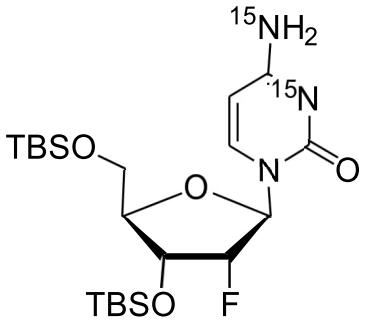
^{19}F-NMR (377 MHz, $CDCl_3$): δ/ppm = -202.7 – -202.4 (m, C2'-*F*).

^{15}N-NMR (41 MHz, $CDCl_3$): *δ* (ppm) = -224.9.

HRMS (ESI): calc. for $C_{21}H_{40}FN^{15}NO_5Si_2^+$ $[M+H]^+$: 476.24246, found: 476.24280.

The data are consistent with literature. [3]

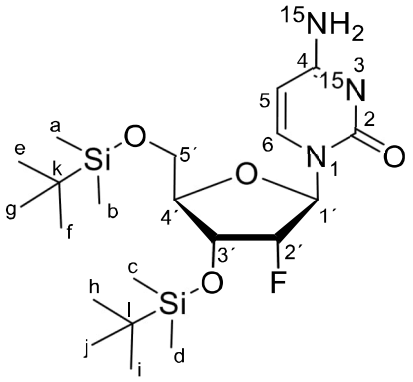
3',5'-bis-*O*-[tert-Butyl(dimethyl)silyl]-2'-deoxy-2'-(*R*)-fluoro-(*N*^{3},*N*^{4}-$^{15}N_2$)-cytidine (13)



13

1,2,4-triazole (0.718 g, 10.4 mmol, 9.9 equiv.) was dissolved in MeCN (20 mL), cooled to 0 °C and stirred at 0 °C for 10 min. $POCl_3$ (0.20 mL, 2.15 mmol, 2.1 equiv.) was slowly added dropwise and the reaction mixture was stirred for additional 10 min at 0 °C. Afterwards, NEt_3 (1.4 mL, 10.0 mmol, 9.5 equiv.) was added dropwise and the mixture was stirred for 20 min at 0 °C. Next, a solution of compound **12** (0.500 g, 1.05 mmol, 1.0 equiv.) in MeCN (7.0 mL) was slowly added dropwise. The reaction mixture was stirred for 16 h at 60 °C. Afterwards, the reaction was quenched by addition of an aqueous, saturated $NaHCO_3$ solution (100 mL), MeCN was removed under reduced pressure and the solution was extracted with CH_2Cl_2 (3 x 100 mL). The combined organic layers were dried over Na_2SO_4 and solvents were removed *in vacuo*. The triazole derivative, which was obtained as a yellow oil, was dissolved in 1,4-dioxane (2.1 mL) and combined dropwise with a solution of $^{15}NH_3$ (3 N, 2.1 mL). The reaction mixture was stirred for 16 h at 30 °C and subsequently stopped by addition of an aqueous, saturated NH_4Cl solution (35 mL) and extracted with CH_2Cl_2 (3 x 100 mL). The

combined organic layers were washed with an aqueous, saturated NaCl solution (100 mL) and dried over Na_2SO_4. Solvents were removed *in vacuo* and the resulting yellowish oil was purified by flash chromatography (SiO_2, CH_2Cl_2 → CH_2Cl_2/MeOH = 99:1 → CH_2Cl_2/MeOH = 96:4) to yield compound **13** as a colourless foam (0.289 g, 0.608 mmol, 58%).



R_f = 0.28 (SiO_2, DCM/MeOH = 99:1).

^{1}H-NMR (500 MHz, $CDCl_3$): δ/ppm = 7.92 (d, ^3J = 8.2 Hz, 1H, C6-H), 6.06 (dd,$^3J_{H\text{-}F}$ = 15.2 Hz, ^3J = 1.9 Hz, 1H, C1'-H), 5.68 (d, ^3J = 8.2 Hz, 1H, C5-H), 4.77 (ddd, $^2J_{H\text{-}F}$ = 52.4 Hz, ^3J = 4.2 Hz, ^3J = 1.9 Hz, 1H, C2'-H), 4.29 (ddd, $^3J_{H\text{-}F}$ = 18.7 Hz, ^3J = 7.1 Hz, ^3J = 4.2 Hz, 1H, C3'-H), 4.11–4.02 (m, 2H, C4'-H, C5'-H), 3.78 (dd, ^2J = 11.7 Hz, ^3J = 1.6 Hz, 1H, C5'-H), 0.95–0.84 (m, 18H, $C_{e\text{-}g}$–H, $C_{h\text{-}j}$-H), 0.16–0.07 (m, 12H, $C_{a\text{-}d}$-H).

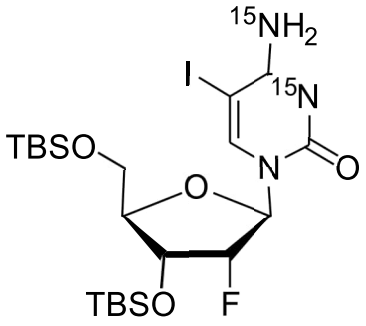
^{13}C-NMR (126 MHz, $CDCl_3$): δ/ppm = 162.7 (d, $^1J_{C\text{-}N}$ = 9.1 Hz, C4), 143.8 (d, $^1J_{C\text{-}N}$ = 18.1 Hz, C2), 141.9 (C6), 101.7 (C5), 93.4 (d, $^1J_{C\text{-}F}$ = 193.5 Hz, C2'), 89.0 (d, $^2J_{C\text{-}F}$ = 33.8 Hz, C1'), 83.7 (C4'), 68.4 (d, $^2J_{C\text{-}F}$ = 16.2 Hz, C3'), 60.7 (C5'), 26.3 ($C_{e\text{-}g}$)*, 25.9 ($C_{h\text{-}j}$)*,18.8 (C_k)**, 18.4 (C_l)**, -4.3 (C_a)***, -4.7 (C C_b)***, -5.0 (C_c)***, -5.1 (C_d)***.

^{19}F-NMR (126 MHz, $CDCl_3$): δ/ppm = -201.7 (dt, $^2J_{H\text{-}F}$ = 35.8 Hz, $^3J_{H\text{-}F}$ = 17.7 Hz, C2'-*F*).

^{15}N-NMR (41 MHz, $CDCl_3$): *δ* (ppm) = -247.3, -283.6.

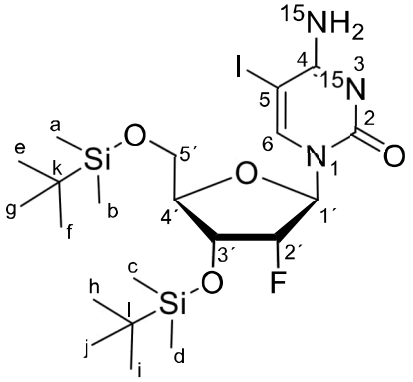
HRMS (ESI): calc. for $C_{21}H_{41}FN^{15}N_2O_4Si_2^+$ $[M+H]^+$: 476.25548, found: 476.25568.

3',5'-bis-*O*-(tert-Butyl(dimethyl)silyl)-5-iodo-2'-deoxy-2'-(*R*)-fluoro-(N^3,N^4-$^{15}N_2$)-cytidine (14)



14

Compound **13** was dissolved in MeCN (35 mL) and CAN (0.540 g, 0.985 mmol, 2.2 equiv.) and iodine (0.250 g, 0.985 mmol, 2.2 equiv.) were added. The reaction mixture was stirred for 3 h at 60 °C and subsequently stopped by addition of an aqueous, saturated $NaHCO_3$ solution (100 mL) and an aqueous, saturated $Na_2S_2O_3$ solution (50 mL). MeCN was removed *in vacuo* and the mixture was extracted with CH_2Cl_2 (3 x 150 mL). The combined organic layers were dried over Na_2SO_4 and solvents were removed *in vacuo*. The resulting yellowish oil was purified by flash chromatography (SiO_2, CH_2Cl_2 → CH_2Cl_2/MeOH = 99:1 → CH_2Cl_2/ MeOH = 98:2) and compound **14** was obtained as a yellow solid (0.251 g, 0.417 mmol, 93%).



R_f = 0.50 (SiO_2, CH_2Cl_2/MeOH = 99:1).

^{1}H-NMR (500 MHz, $CDCl_3$): δ/ppm = 8.15 (s, 1H, C6-H), 6.03 (dd, $^3J_{H\text{-}F}$ = 15.0 Hz, ^3J = 2.5 Hz, 1H, C1'-H), 4.88 (dt, 1H, C2'-H), 4.24 (ddd, $^3J_{H\text{-}F}$ = 16.4 Hz, ^3J = 6.6 Hz, ^3J = 4.5 Hz, 1H, C3'-H), 4.12 (d, ^3J = 7.0 Hz, 1H, C5'-H), 4.05 (dd, $^3J_{H\text{-}F}$ = 11.9 Hz, ^3J = 1.7 Hz, 1H, C4'-H), 3.79 (dd, ^2J = 12.0 Hz, ^3J = 2.2 Hz, 1H, C5'-H), 0.96 (s, 9H, $C_{e\text{-}g}$-H), 0.91 (s, 9H, $C_{h\text{-}j}$-H), 0.17 (2 x s, 6H, C_a-H, C_b-H)*, 0.13 (s, 3H, C_c-H)*, 0.11 (s, 3H, C_d-H)*.

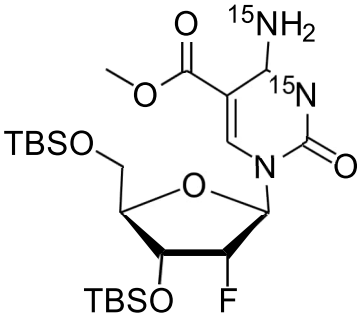
^{13}C-NMR (126 MHz, $CDCl_3$): δ/ppm = 165.0 (C4), 150.8 (C2), 147.1 (C6), 92.0 (d, $^1J_{C-F}$ = 200.8 Hz, C2'), 88.3 (d, $^2J_{C-F}$ = 34.3 Hz, C1'), 84.2 (C4'), 68.2 (d, $^2J_{C-F}$ = 15.1 Hz, C3'), 60.5 (C5'), 55.7 (C5), 25.7 (C_{e-g})*, 25.0 (C_{h-j})*, 18.0 (C_k)**, 17.4 (C_l)**, -5.26 (C_a)***, -5.38 (C_b)***, -5.51 (C_c)***, -5.75 (C_d)***.

^{19}F-NMR (126 MHz, $CDCl_3$): δ/ppm = -203.1 (d, $^2J_{H-F}$ = 47.9 Hz, C2'-*F*).

^{15}N-NMR (41 MHz, $CDCl_3$): *δ* (ppm) = -248.6, -267.1.

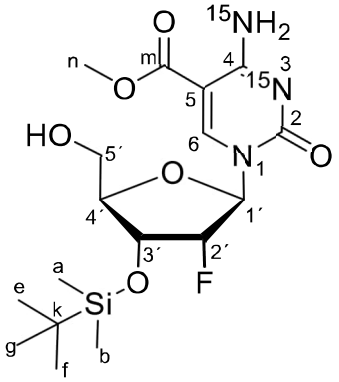
HRMS (ESI): calc. for $C_{21}H_{40}FIN^{15}N_2O_4Si_2^+$ $[M+H]^+$: 602.15213, found: 602.15288.

3',5'-bis-*O*-(tert-Butyl(dimethyl)silyl)-5-(methylcarboxy)-2'-deoxy-2'-(*R*)-fluoro-(*N*^{3},*N*^{4}-$^{15}N_2$)-cytidine (15)



15

The reaction vessel of a high-pressure CO-autoclave was charged with **14** (50.0 mg, 83.1 µmol, 1.0 equiv.) and $[Pd(MeCN)_2Cl_2]$ (11.0 mg, 4.24 µmol, 5.0 Mol%). The solids were taken up in MeOH (1 mL). The reactor was purged with CO (3.5 bar) to test for leakage. Subsequently, the reactor was purged two more times with CO (3.5 bar). DIPEA (28.0 µL, 0.16 mmol, 2.0 equiv.) was added and the reaction mixture was stirred for 20 h at 65 °C at 3.5 bar CO. The reaction mixture was filtered to remove residual solids and the filtrate was concentrated *in vacuo*. Purification by flash chromatography (SiO_2, CH_2Cl_2 → CH_2Cl_2/MeOH = 99:1 → CH_2Cl_2/MeOH = 98:2) yielded carboxylated compound **15** (33.0 mg, 61.8 µmol, 74%) as a colourless foam.



R_f = 0.57 (SiO_2, CH_2Cl_2/MeOH = 10:1).

^{1}H-NMR (599 MHz, $CDCl_3$): δ/ppm = 9.15 (s, 1H, C6-H), 8.71 (d, ^2J = 93.6 Hz 2H, C4-NH_2), 5.91 (dd, $^3J_{H–F}$ = 16.0 Hz, 1H, C1'-H), 5.20 (ddd, $^2J_{H–F}$ = 51.9 Hz, ^3J = 4.4 Hz, 1H, C2'-H), 4.45 (ddd, $^3J_{H–F}$ = 20.1 Hz, ^3J = 7.9 Hz, ^3J = 4.2 Hz 1H, C3'-H), 4.16 (dd, ^3J = 20.2 Hz, ^3J = 7.7 Hz,1H, C4'-H), .3.90 (s, 3H, C_n-H_3), 3.71-3.63 (m 1H, C5'-H_a)*, 3.14-3.07 (m, 1H, C5'-H_b)*, 0.91 (s, 9H, C_{e-g}-H_3), 0.15 (s, 3H, C_a)**, 0.13 (s, 3H, C_b)**.

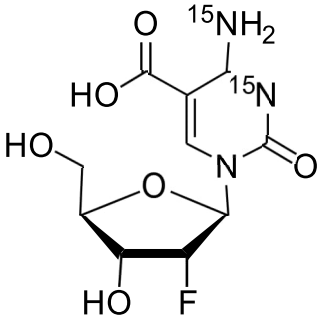
^{13}C-NMR, HSQC, HMBC (126 MHz, $CDCl_3$): δ/ppm = 163.0, 159.0, 162.4, 150.1, 90.2, 95.6, 92.0, 89.3, 84.7, 68.3, 59.8, 53.6, 64.2, 42.6, 25.7, 18.0, -5.26, -5.38.

^{15}N-NMR (41 MHz, $CDCl_3$): *δ* (ppm) = -234.9, -274.3.

HRMS (ESI): calc. for $C_{17}H_{28}FN^{15}N_2O_6Si$ $[M+Na]^+$ 442.1564, found: 442.1561.

IR (ATR): v/cm^{-1} = 3359 (w), 3255 (w), 3104 (w), 2953 (w), 2930 (w), 2892 (w), 2858 (w), 1760 (m), 1713 (s), 1694 (s), 1642 (w), 1614 (m), 1453 (s), 1406 (w), 1391 (w), 1328 (w), 1298 (w), 1254 (s), 1230 (w), 1193 (w), 1162 (w), 1122 (m), 1084 (s), 1069 (s), 1005 (w), 983 (w), 938 (w), 910 (w), 863 (s), 840 (vs), 815 (s), 780 (vs), 758 (s), 672 (m).

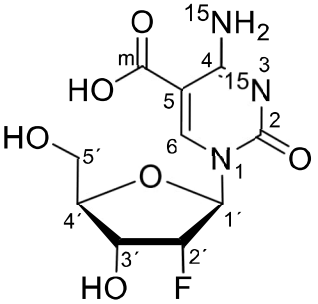
5-(Carboxy)-2'-deoxy-2'-(*R*)-fluoro-(*N*^{3},*N*^{4}-$^{15}N_2$)-cytidine (3)



3

Compound **15** (30.0 mg, 71.5 μmol, 1.0 equiv.) was dissolved in H_2O/MeCN = 3:5 (3 mL/5 mL), LiOH (21.0 mg, 0.86 mmol, 12.0 equiv.) was added and the reaction mixture was stirred for 20 h at room temperature. HCl (2 M, 10 mL) was then added until the solution showed pH = 3 and the mixture was concentrated *in vacuo*. The obtained nucleoside was used without further processing for the final deprotection.

For this, the nucleoside (71.5 μmol, 1.0 equiv.) was dissolved in EtOAc (1.0 mL) and HF • Pyridin (70%, 21.0 mg, 28.0 μL, 1.07 mmol, 15.0 equiv.) was added. The reaction mixture was stirred for 20 h at room temperature, then cooled to 0 °C, combined with TMS • OMe (0.23 g, 0.30 mL, 2.29 mmol, 32.0 equiv.) and stirred for 1 h at 0 °C. The obtained nucleoside was purified by semi-preparative HPLC (MeCN/H_2O, MeCN: 0 → 20%, 45 min, retention time: 10 min) yielded nucleoside **3** (2.44 mg, 8.38 μmol, 12%) after lyophilisation as a colourless powder.



Retention time = 10 min (C_{18}-SiO_2, MeCN/H_2O, MeCN: 0 → 20%, 45 min).

^{1}H-NMR (800 MHz, D_2O): δ/ppm = 8.71 (s, 1H, C6-H), 6.10 (d, $^3J_{H–F}$ = 18.5 Hz, 1H, C1'-H), 5.21 (dd, $^2J_{H–F}$ = 52.5 Hz, ^3J = 4.6 Hz, 1H, C2'-H), 4.41 (ddd, $^3J_{H–F}$ = 22.6 Hz, ^3J = 8.8 Hz,

^{3}J = 4.4 Hz 1H, C3'-H), 4.22 (d, ^{3}J = 9.0 Hz 1H, C4'-H), 4.09 (dd, ^{2}J = 13.2 Hz, ^{3}J = 2.4 Hz, 1H, C5'-H_a)*, 3.90 (dd, ^{2}J = 13.0 Hz, ^{3}J = 4.1 Hz, 1H, C5'-H_b)*.

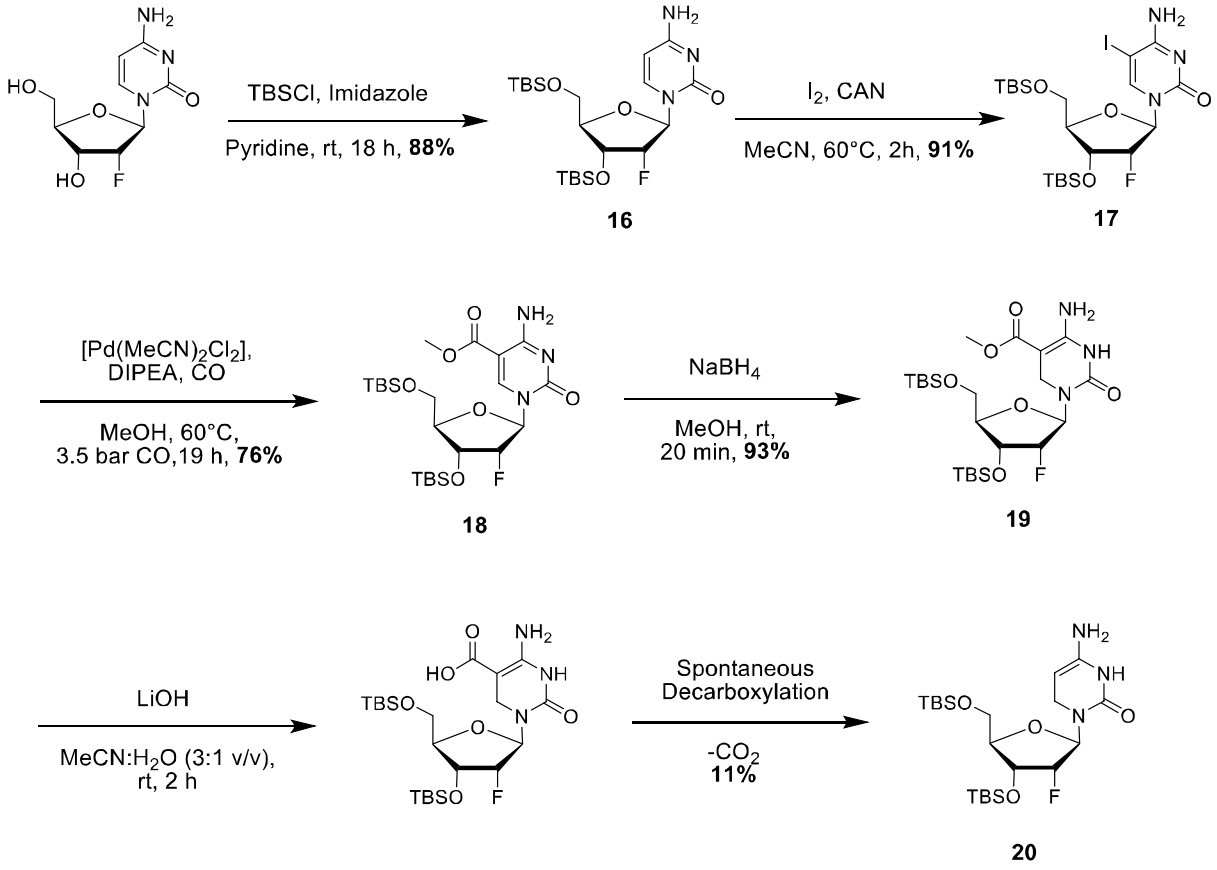
^{13}C-NMR, HSQC, HMBC (101 MHz, D_2O): δ/ppm = 168.8 (C_m), 162.2 (C-4), 152.6 (C-2), 147.6 (C-6), 93.3, (C2'-H), 89.7, (C1'-H), 82.3 (C4'-H), 67.3, (C3'-H), 59.5 (C5'-H).

^{19}F-NMR (377 MHz, D_2O): δ/ppm = −201.1 - −201.5 (m, C2'-*F*).

HRMS (ESI): calc. $C_{10}H_{13}FN^{15}N_2O_6^+$ $[M+H]^+$ 292.0724, found: 292.0721.

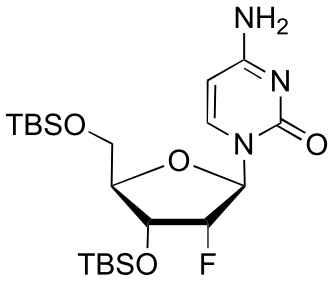
IR (ATR): v/cm^{-1} = 3334 (m), 3206 (m), 2928 (w), 2892 (w), 2857 (w), 2227 (w), 1647 (vs), 1493 (s), 1408 (m), 1361 (m), 1295 (m), 1252 (m), 1097 (s), 1060 (s), 986 (w), 950 (w), 917 (w), 833 (m), 813 (w), 781 (m), 703 (w), 672 (w).

In vitro decarboxylation experiment



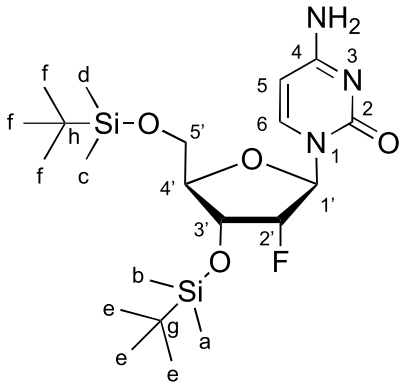
Scheme SI-2: Synthesis pathway towards 2'-F-5,6-dihydro-dC (**20**).

O-3',5'-bis(*tert*-Butyldimethylsilyl)-2'-deoxy-2'-(R)-fluoro-cytidine (16)



16

2'-deoxy-2'-fluoro-cytidine (1 g, 4.1 mmol, 1 equiv.) was dissolved in pyridine (0.5 M) and imidazole (1.25 g, 18.4 mmol, 4.5 equiv.) and TBSCl (1.85 g, 12.3 mmol, 3.0 equiv.) were added. The mixture was stirred for 17 h at room temperature. Pyridine was removed under reduced pressure; the resulting residue was taken up in an aqueous $NaHCO_3$ solution (100 mL) and extracted with CH_2Cl_2 (3 x 150 mL). The combined organic layers were washed with an aqueous, saturated NaCl solution (100 mL) and dried over Na_2SO_4. Solvents were removed *in vacuo* and the resulting yellow oill was coevaporated with toluene (3 x 50 mL). The TBS-protected nucleoside **16** was yielded as a colourless, waxy solid (1.7 g, 3.6 mmol, 88%).



TLC: R_f = 0.60 (DCM/MeOH 10:1)

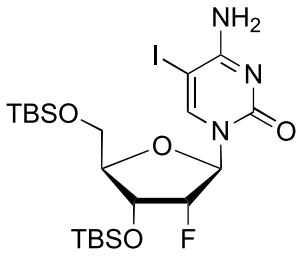
^{1}H-NMR (400 MHz, $CDCl_3$, ppm): δ = 9.20 (s, 1H, NH_2), 7.97 (s, 1H, 6-H), 5.96 (dd, $^3J_{H\text{-}F}$ = 16,5 Hz, ^3J = 1.8 Hz, 1H, 1'-H), 5.69 (s, 1H, NH_2), 4.79 (ddd, $^2J_{H\text{-}F}$ = 52,4 Hz, ^3J = 4.3 Hz, ^3J = 1.8 Hz, 1H, 2'-H), 4.17 (ddd, $^3J_{H\text{-}F}$ = 18.8 Hz, ^3J = 7.3 Hz, ^3J = 4.2 Hz, 1H, 3'-H), 4.05 - 4.02 (m, 2H, 4'-H, 5'-H), 3.76 (dd, ^2J = 11.9 Hz, ^3J = 2.4 Hz, 1H, 5'-H)*, 0.93 (s, 9H, Ce-H)*, 0.86 (s, 9H, Cf-H)**, 0.14 (s, 3H, Ca-H)***, 0.13 (s, 3H, Cb-H)***, 0.08 (s, 3H, Cc-H)***, 0.06 (s, 3H, Cd-H)***.

^{13}C-NMR (75 MHz, $CDCl_3$, ppm): δ = 164.1, 154.6, 146.0, 93.5 (d, *J* = 192.8 Hz), 89.0 (d, *J* = 34.1 Hz), 83.7, 68.6 (d, *J* = 16.7 Hz), 60.9, 57.2, 26.4, 25.6, 18.7, 18.1, −4.6, −4.9, −5.0, −5.1.

^{19}F-NMR (377 MHz, $CDCl_3$, ppm): -200.9 (ddd, $^2J_{H-F}$ = 52,5 Hz, $^3J_{H-F}$ = 24.1 Hz, $^3J_{H-F}$ = 16,3 Hz)

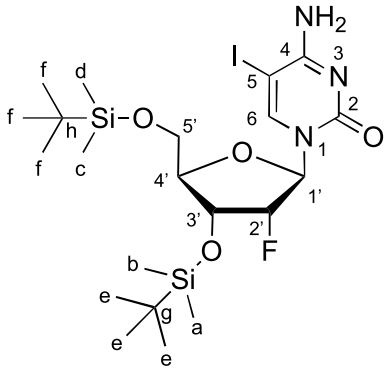
HRMS (ESI) calc. $C_{21}H_{41}FN_3O_4Si_2^+$ $[M+H]^+$, 474.26141, found: 474.26160.
calc. $C_{21}H_{39}FN_3O_4Si_2^-$ $[M-H]^-$, 472.24686, found: 472.24715

O-3',5'-bis(*tert*-Butyldimethylsilyl)-5-iodo-2'-deoxy-2'-(R)-fluoro-cytidine (17)



17

Compound **16** (1.1 g, 2.3 mmol, 1 equiv.) was dissolved in MeCN (100 mL) and CAN (2.8 g, 5.1 mmol, 2.2 equiv.) and iodine (1.3 g, 5.1 mmol, 2.2 equiv.) were added. The reaction mixture was stirred for 3 h at 60 °C and subsequently stopped by addition of an aqueous, saturated $NaHCO_3$ solution (100 mL) and an aqueous, saturated $Na_2S_2O_3$ solution (50 mL). MeCN was removed *in vacuo* and the mixture was extracted with CH_2Cl_2 (3 x 150 mL). The combined organic layers were dried over Na_2SO_4 and solvents were removed *in vacuo*. The resulting yellowish oil was purified by flash chromatography (SiO_2, CH_2Cl_2 → CH_2Cl_2/MeOH = 99:1 → CH_2Cl_2/ MeOH = 98:2) and compound **17** was obtained as a pale yellow solid (0.99 g, 2.1 mmol, 91%).



TLC: R_f = 0.57 (DCM/MeOH 10:1)

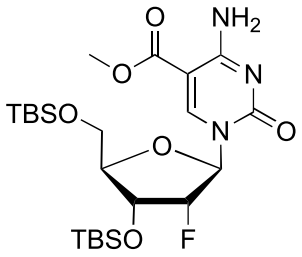
^{1}H-NMR (400 MHz, $CDCl_3$, ppm): δ = 8.36 (s, 1H, NH_2), 7.97 (d, ^3J = 7.4 Hz, 1H, 6-H), 6.19 (s, 1H, NH_2), 6.01 (d, $^3J_{H-F}$ = 16.5 Hz, 3 J = 1.8 Hz, 1H, C1'-H), 5,78 (d, ^3J = 7.5 Hz, 1H, C5-H), 4,76 (dd, $^2J_{H-F}$ = 52,2 Hz, ^3J = 3,8 Hz, 1H, C2'-H), 4.21 (ddd, $^3J_{H-F}$ = 23.3 Hz, ^3J = 8.5 Hz, ^3J = 3.9 Hz, 1H, C3'-H), 4,08 (dd, ^3J = 11.9 Hz, ^3J = 1.8 Hz, 1H, C4'-H), 4,05 (dd, ^2J = 8,8 Hz, ^3J = 2,0 Hz, 1H, C5'-H)*, 3,78 (dd, ^2J = 11.8 Hz, ^3J = 1.6 Hz, 1H, C5'-H)*, 0.92 (s, 9H, Ce-H)**, 0.87 (s, 9H, Cf-H)**, 0.11 (s, 3H, Ca-H)***, 0.10 (s, 3H, Cb-H)***, 0.08 (s, 3H, Cc-H)***, 0.06 (s, 3H, Cd-H)***

^{13}C-NMR (75 MHz, $CDCl_3$, ppm): δ = 166.3, 155,7, 140.4, 94,9, 93.5 (d, J = 191.1 Hz), 88.7 (d, J = 33.4 Hz), 82,6, 67.8 (d, J = 16.7 Hz), 60.2, 57.2, 26.0, 25.6, 18.4, 18.0, −4.6, −5.0, −5.5, −5.5.

^{19}F-NMR (377 MHz, $CDCl_3$, ppm): δ = -201.7 (dt, $^2J_{H-F}$ = 52.2 Hz, $^3J_{H-F}$ = 17.2 Hz)

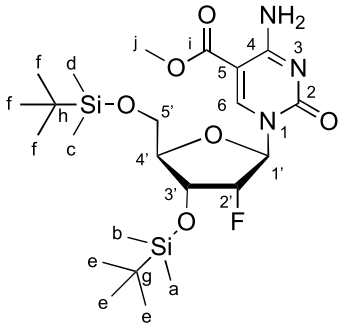
HRMS (ESI) calc. $C_{21}H_{40}FIN_3O_4Si_2^+$ $[M+H]^+$, 600.15806, found: 600.15848.
calc. $C_{21}H_{38}FIN_3O_4Si_2^-$ $[M-H]^-$, 598.14350, found: 598.14399

3'-5'-*O*-bis(*tert*-Butyldimethylsilyl)-5-methoxycarbonyl-2'-desoxy-2'-(R)-fluoro-cytidine (18)



18

2.00 g of compound **17** (3.03 mmol, 1.0 equiv.) and 43 mg $[Pd(MeCN)_2Cl_2]$ (165 µmol, 0.05 equiv.) were placed in a high-pressure glass autoclave. To test its tightness, the autoclave was filled with CO gas to 3.5 bar and left for 10 min. Subsequently, the CO gas was released, 1.10 mL DIPEA (6.60 mmol, 2.0 equiv.) and 36 mL methanol (0.9 M) were added. The autoclave was flushed twice with CO gas and finally filled with CO gas to 3.5 bar. The suspension was heated to 60 °C for 18 h. After release of the CO gas, removal of the solvent *in vacuo* and purification by column chromatography (dryload, DCM/MeOH 1% → 5%), 1.3 g of compound 22 (2.3 mmol, **76%**) were obtained as a pale yellow solid.



TLC R_f = 0.33 (5 % MeOH in DCM)

^{1}H NMR (400 MHz, $CDCl_3$) δ = **8.49** (s, 1H, C6-H) **8.17** (s, 1H, 1·NH_2), **7.65** (s, 1H, 1·NH_2), **5.92** (dd, 1H, $^3J_{H-F}$ = 17.0 Hz, ^3J = 2.2 Hz, C1'-H), **4.97** (ddd, 1H, $^2J_{H-F}$ = 52.1 Hz, ^3J = 4.1 Hz, ^3J = 2.1 Hz, C2'-H), **4.21** (ddd, 1H, $^3J_{H-F}$ = 18.1 Hz, ^3J = 7.3 Hz, ^3J =3.7 Hz, C3'-H), **4.12-4.08** (m, 1H, C4'-H), **4.01** (dd, 1H, ^2J = 12.1 Hz, ^3J = 2.9 Hz, C5'-H)*, **3.83** (s, 3H, Cj-H), **3.79** (dd, 1H, ^2J = 12.0 Hz, ^3J = 3.8 Hz,

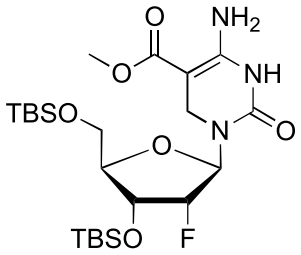
C5'-H)*, **0.90** (s, 9H, Ce-H)** **0.89** (s, 9H, Cf-H)**, **0.1** (s, 3H, Ca-H)***, **0.1** (s, 3H, Cb-H)***, **0.09** (s, 3H, Cc-H)***, **0.08**(s, 3H, Cd-H)***.

^{13}C NMR (75 MHz, $CDCl_3$) δ = 165.1, 163.6, 153.6, 148.6, 95.5, 92.9 (d, $^1J_{C-F}$ = 193.3 Hz), 90.8 (d, J = 34.4 Hz), 84.7, 69.5 (d, J = 16.4 Hz), 61.9, 52.5, 26.3 (3C), 26.0 (3C), 18.8, 18.4, −4.3, −4.5, −5.0, −5.2.

^{19}F-NMR (377 MHz, $CDCl_3$): δ/ppm = -201.2 (dt, $^2J_{F2'-H2'}$ = 51.5 Hz, $^3J_{F-H}$ = 18.5 Hz)

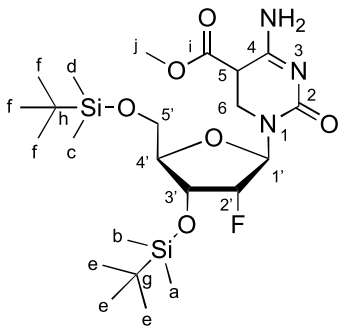
HRMS (ESI): calc. $C_{23}H_{43}FN_3O_6Si_2^+$ $[M+H]^+$, 532.26689, found: 532.26712.

3'-5'-*O*-bis(*tert*-Butyldimethylsilyl)-5-methoxycarbonyl-2'-desoxy-2'-(R)-fluoro-5,6-dihydrocytidine (19)



19

To 3´,5´-(tert-Butyl-dimethylsilyl)-5-carboxymethyl-2´-fluoro-2'-deoxy-cytidine (52.0 mg, 97.32 µmol, 1.0 equiv.) dissolved in methanol (5.0 mL) was added sodium borohydride (3.6 mg, 95.16 µmol, 0.98 equiv.). After stirring for 20 min at room temperature a saturated NH_4Cl solution (33 mL) was added and extracted with EtOAc (33 mL). The organic phase was washed with a saturated NH_4Cl solution (2 · 30 mL), dried over $MgSO_4$ and evaporated to dryness to give the target as an off white solid (50 mg, 93.4 µmol, **96%**).



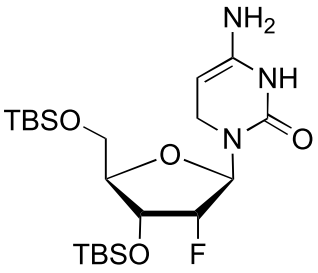
^{1}H NMR (400 MHz, $CDCl_3$) δ = **9.40** (s, 1H, N3-H), **6.44** (s, br, 2H, 2·NH_2), **6.06** (dd, 1H, $^{3}J_{H\text{-}F}$ = 18.3 Hz, ^{3}J = 4.0 Hz, C1'-H), **4.85** (dt, 1H, $^{2}J_{H\text{-}F}$ = 53.6 Hz, ^{3}J = 4.6 Hz, C2'-H), **4.29** (dt, 1H, $^{3}J_{H\text{-}F}$ = 12.8 Hz, ^{3}J = 5.2 Hz, C3'-H), **4.09** (d, 1H, J = 12.3 Hz, 1·C6-H), **3.91-3.81** (m, 3H, 1·C6-H, C4'-H, 1·C'5-H), **3.71** (dd, 1H, ^{2}J = 11.6 Hz, ^{3}J = 2.7 Hz, 1·C5'-H), **3.65** (s, 3H, Cg-H), **0.91** (s, 9H, Ce-H)* **0.90** (s, 9H, Cf-H)*, **0.12** (s, 3H, Ca-H)**, **0.11** (s, 3H, Cb-H)**, **0.07** (s, 3H, Cc-H)**, **0.06** (s, 3H, Cd-H)**.

^{13}C NMR (75 MHz, $CDCl_3$) δ = 168.1, 154.9, 90.4 (d, $^{1}J_{C\text{-}F}$ = 191.6 Hz), 86.5 (d, J = 32.9 Hz), 83.3, 70.2 (J = 15.4), 62.1, 39.7, 26.3 (3C), 26.0 (3C), 18.5, 18.7, -4.4, -4.7, -5.1, -5.3.

^{19}F NMR (377 MHz, $CDCl_3$) δ = -206.3 (ddd, $^{2}J_{F2'\text{-}H2'}$ = 52.8 Hz, $^{3}J_{F\text{-}H'}$ = 18.3 Hz, $^{3}J_{F\text{-}H'}$ = 13.4 Hz)

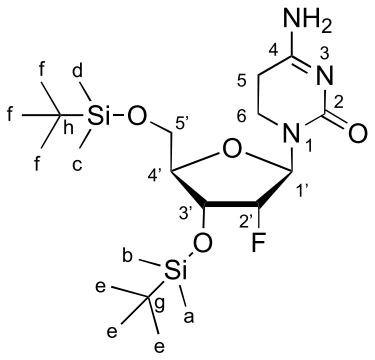
HRMS (ESI) calc. $C_{23}H_{45}FN_3O_6Si_2^{+}$ $[M+H]^{+}$, 534.28254, found: 534.28264.
calc. $C_{23}H_{43}FN_3O_6Si_2^{-}$ $[M-H]^{-}$, 532.26799, found: 532.26838

3'-5'-*O*-bis(*tert*-Butyldimethylsilyl)-2'-desoxy-2'-(R)-fluoro -5,6-dihydrocytidine (20)



20

To a solution of **18** (200 mg, 0.4 mmol, 1.0 equiv.) in an acetonitrile/water mixture (17 mL/5 mL) LiOH (117 mg, 4.9 mmol, 12.2 equiv.) was added and the resulting solution was stirred at room temperature for 2 h. After addition of a saturated $NaHCO_3$-solution (40 mL) and acidification with 2 M hydrochloric acid (20 mL) the aqueous phase was extracted with diethyl ether (3 · 40 mL), dried over $MgSO_4$ and evaporated to dryness. After column chromatography (DCM/MeOH 50:1 → 5:1) the product was obtained as a colorless, waxy solid (22 mg, 0.04 mmol, **11%**).



^{1}H NMR (400 MHz, $CDCl_3$) δ = **5.99** (dd, 1H, $^3J_{H\,F}$ = 19.0 Hz, ^3J = 3.4 Hz, C1'-H), **4.75** (dt, 1H, $^2J_{H-F}$ = 53.7 Hz, ^3J = 4.1 Hz, C2'-H), **4.23** (dt, 1H, $^3J_{H-F}$ = 15.1 Hz, ^3J = 5.5 Hz, C3'-H), **3.89-3.83** (m, 2H, 1·C5-H, C4'-H), **3.70** (dd, 1H, ^2J = 12.2 Hz, ^3J = 3.4 Hz, 1·C5'-H), **3.56** (dt, 1H, ^3J = 12.1 Hz, ^3J = 6.0 Hz, 1·C6-H), **3.28-3.18** (m, 1H, 1·C6-H), **2.66-2.50** (m, 1H, 2·C5-H), **0.90** (s, 9H, Ce-H)* **0.88** (s, 9H, Cf-H)*, **0.10** (s, 3H, Ca-H)**, **0.09** (s, 3H, Cb-H)**, **0.06** (s, 3H, Cc-H)**, **0.05**(s, 3H, Cd-H)**.

^{13}C NMR (75 MHz, $CDCl_3$) δ = 166.6, 157.6, 91.8 (d, $^1J_{C-F}$ = 191.0 Hz), 88.0 (d, J = 33.2 Hz), 82.9, 70.1 (J = 17.1), 62.1, 37.9, 26.2 (3C), 26.0 (3C), 18.5, 18.7 -4.4, -4.7, -5.1, -5.2.

^{19}F NMR (377 MHz, $CDCl_3$) δ = -205.7 (dt, $^2J_{F2'-H2'}$ = 52.4 Hz, $^3J_{F-H'}$ = 12.5 Hz)

HRMS (ESI) calc. $C_{21}H_{43}FN_3O_4Si_2^+$ $[M+H]^+$, 476.27706, found: 476.27754.
calc. $C_{21}H_{41}FN_3O_4Si_2^-$ $[M-H]^-$, 474.26251, found: 474.26331.

Supplementary tables

Table SI-1: Compound-dependent LC-MS/MS-parameters used for the analysis of genomic DNA fed with fluorinated nucleosides. *CE: collision energy, CAV: collision cell accelerator voltage, EMV: electron multiplier voltage. The nucleosides were analysed in the positive ($[M+H)]^+$ species) as well as the negative ($[M-H]^-$ species) ion selected reaction monitoring mode (SRM).*

compound	Precursor ion (m/z)	MS1 Resolution	Product Ion (m/z)	MS2 Resolution	Dwell time [ms]	CE (V)	CAV (V)	Polarity
Time segment 1.5-3.7 min								
F-dC	246.09	Wide	112.06	Wide	70	15	3	Positive
[$^{15}N_2$]-F-dC	248.08	Wide	114.04	Wide	70	15	3	Positive
[$^{15}N_2$]-cadC	274.08	Wide	158.03	Wide	40	5	5	Positive
cadC	272.09	Wide	156.04	Wide	40	5	5	Positive
F-hmdC	276.10	Wide	142.06	Wide	50	10	3	Positive
[$^{15}N_2$]-F-hmdC	278.09	Wide	144.06	Wide	50	10	3	Positive
[$^{15}N_2$,D_2]-hmdC	262.12	Wide	146.07	Wide	25	27	1	Positive
hmdC	258.11	Wide	142.06	Wide	25	27	1	Positive
[D_3]-mdC	245.13	Wide	129.09	Wide	50	60	1	Positive
mdC	242.11	Wide	126.07	Wide	50	60	1	Positive
F-cadC	290.08	Wide	156.04	Wide	80	5	5	Positive
[$^{15}N_2$]-F-cadC	292.08	Wide	158.04	Wide	80	5	5	Positive
Time segment 3.7-4.7 min								

[D_3]-F-mdC	263.12	Wide	129.09	Wide	80	15	3	Positive
F-mdC	260.10	Wide	126.07	Wide	80	15	3	Positive
F-cadC	290.08	Wide	156.04	Wide	80	5	5	Positive
[$^{15}N_2$]-F-cadC	292.08	Wide	158.04	Wide	80	5	5	Positive
Time segment 4.7-10 min								
F-fdC	274.08	Wide	140.05	Wide	90	15	3	Positive
[$^{15}N_2$]-F-fdC	276.08	Wide	142.04	Wide	90	15	3	Positive
F-fdU	273.05	Wide	253.05	Wide	30	3	5	Negative
F-dT	259.07	Wide	239.07	Wide	70	3	5	Negative
dT	243.10	Wide	127.05	Wide	20	5	5	Negative
[$^{15}N_2$]-fdC	258.09	Wide	142.04	Wide	30	5	5	Positive
fdC	256.09	Wide	140.05	Wide	30	5	5	Positive
[$^{15}N_2$]-fdC	142.04	Wide	98.04	Wide	20	13	7	Positive
fdC	140.05	Wide	97.04	Wide	20	13	7	Positive

Table SI-2: Compound-dependent LC-MS/MS ranges of the corresponding linear equations.

compound	n (ULOQ)	n (LLOQ)	A/A* (ULOQ)	A/A* (LLOQ)
F-fdC	25.1 fmol	0.390 fmol	0.2267	0.003345
F-dC	800 fmol	3.13 fmol	4.074	0.01512
F-mdC	49.9 fmol	0.780 fmol	0.04174	0.0005833
F-cadC	12.7	0.2	0.240576	0.0217614
F-cadC	202.5	6.3	3.92816	0.1271398

Appendix

NMR Spectra

3',5'-di-*O*-Acetyl-2'-deoxy-2'-(*R*)-fluorouridine (9)

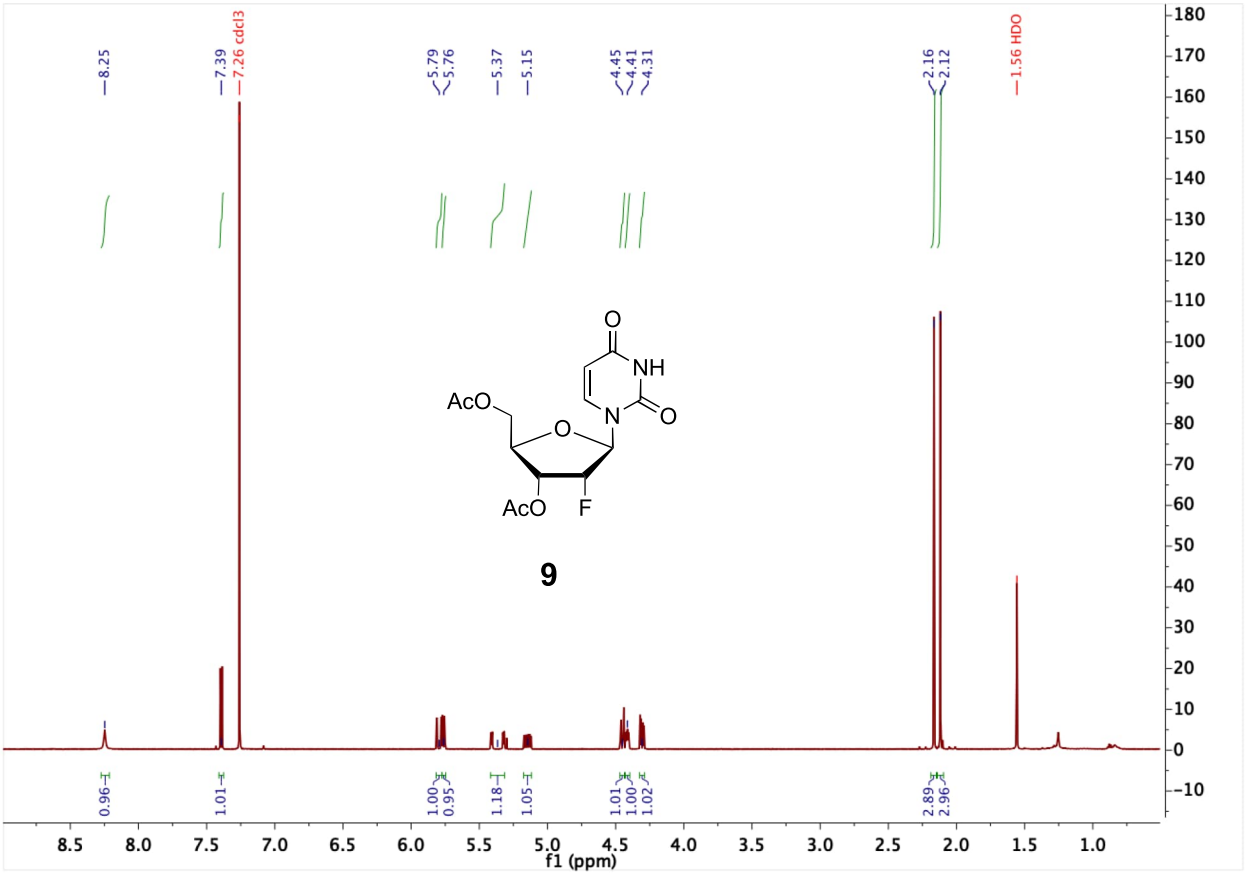


Figure SI-9: ^{1}H-NMR spectrum of compound **9**.

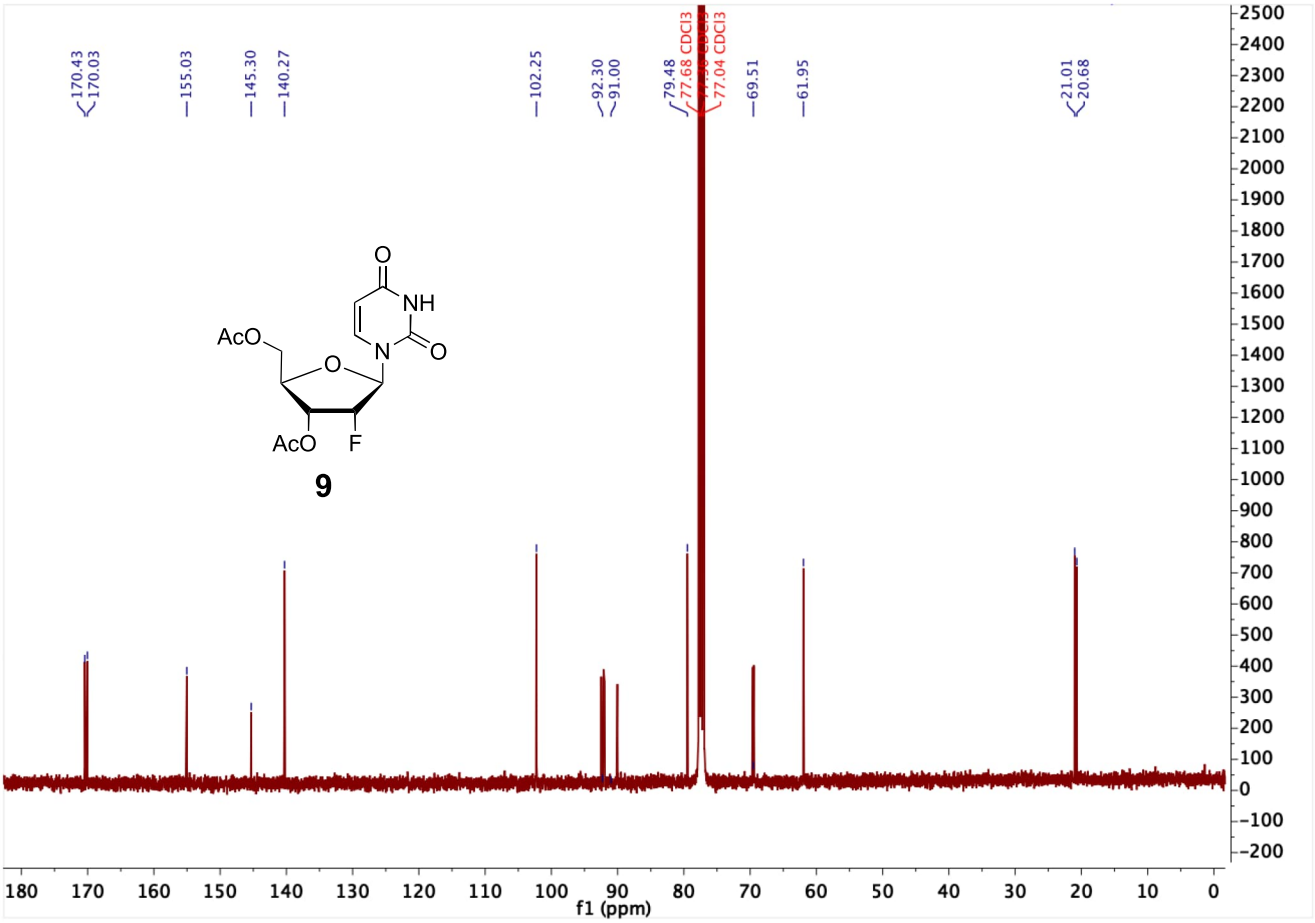


Figure SI-10: ^{13}C-NMR spectrum of compound **9**.

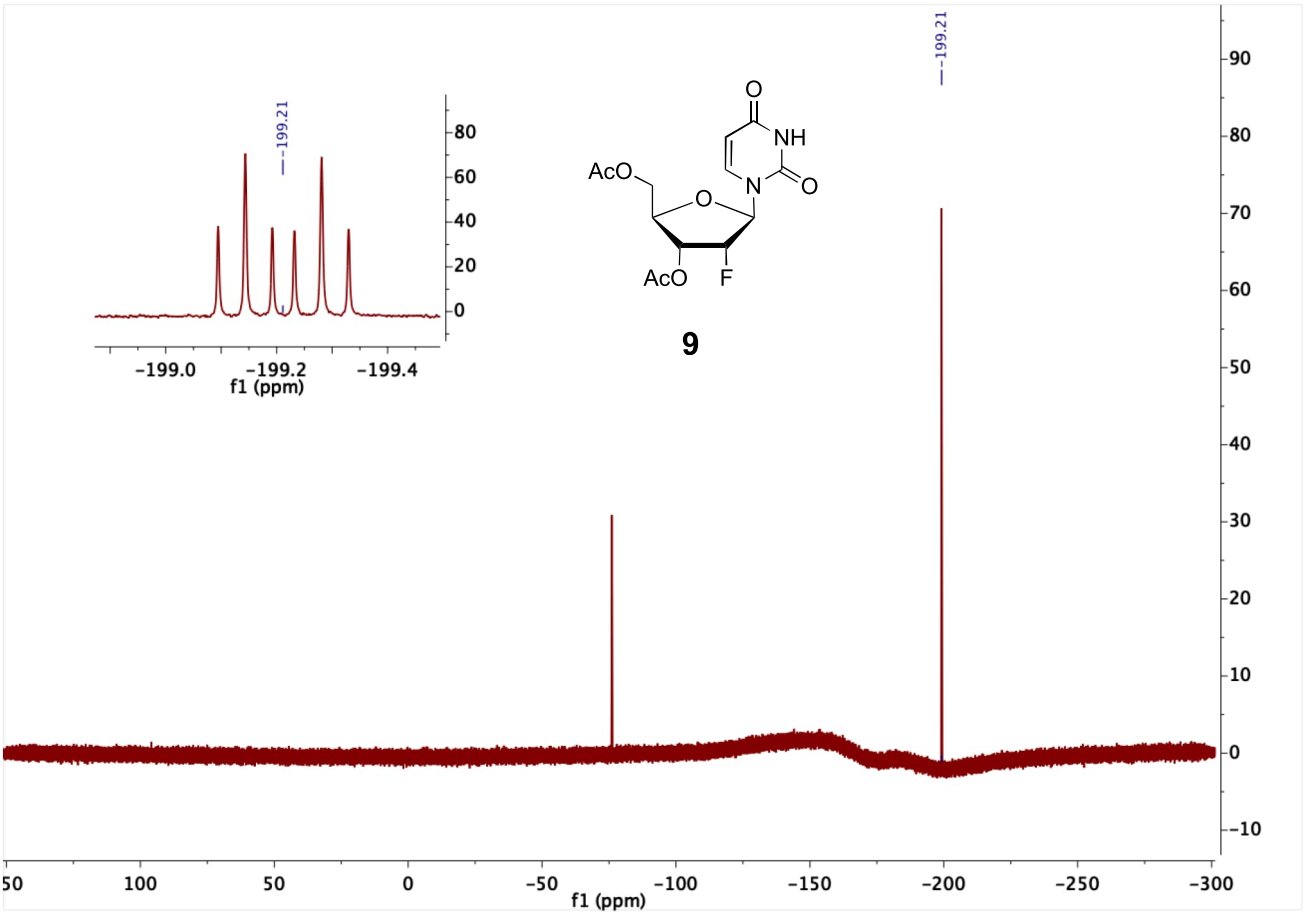


Figure SI-11: ^{19}F-NMR spectrum of compound **9**.

3',5'-di-*O*-Acetyl-2'-deoxy-2'-(*R*)-fluoro-3-nitrouridine (10)

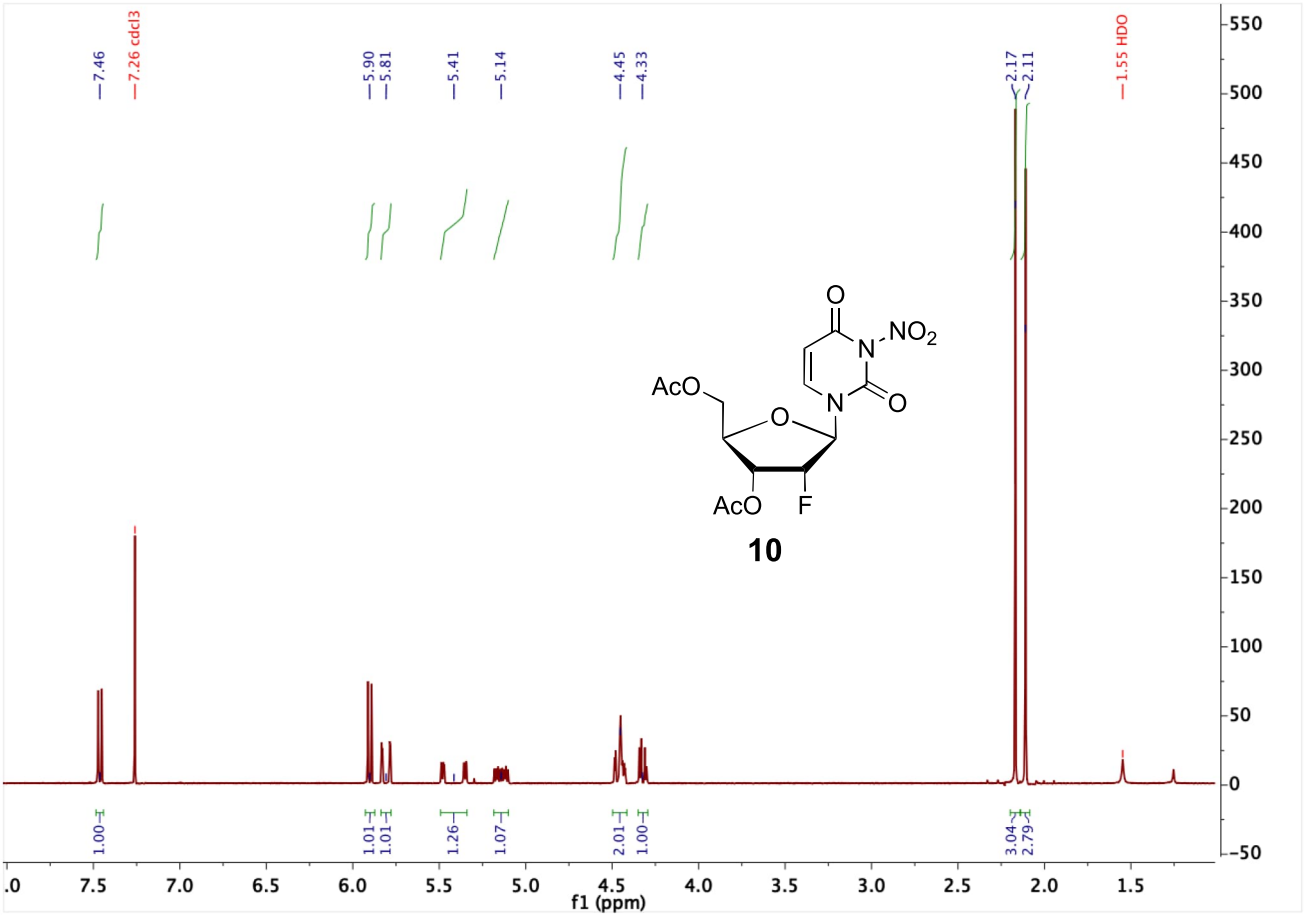


Figure SI-12: ^{1}H-NMR spectrum of compound **10**.

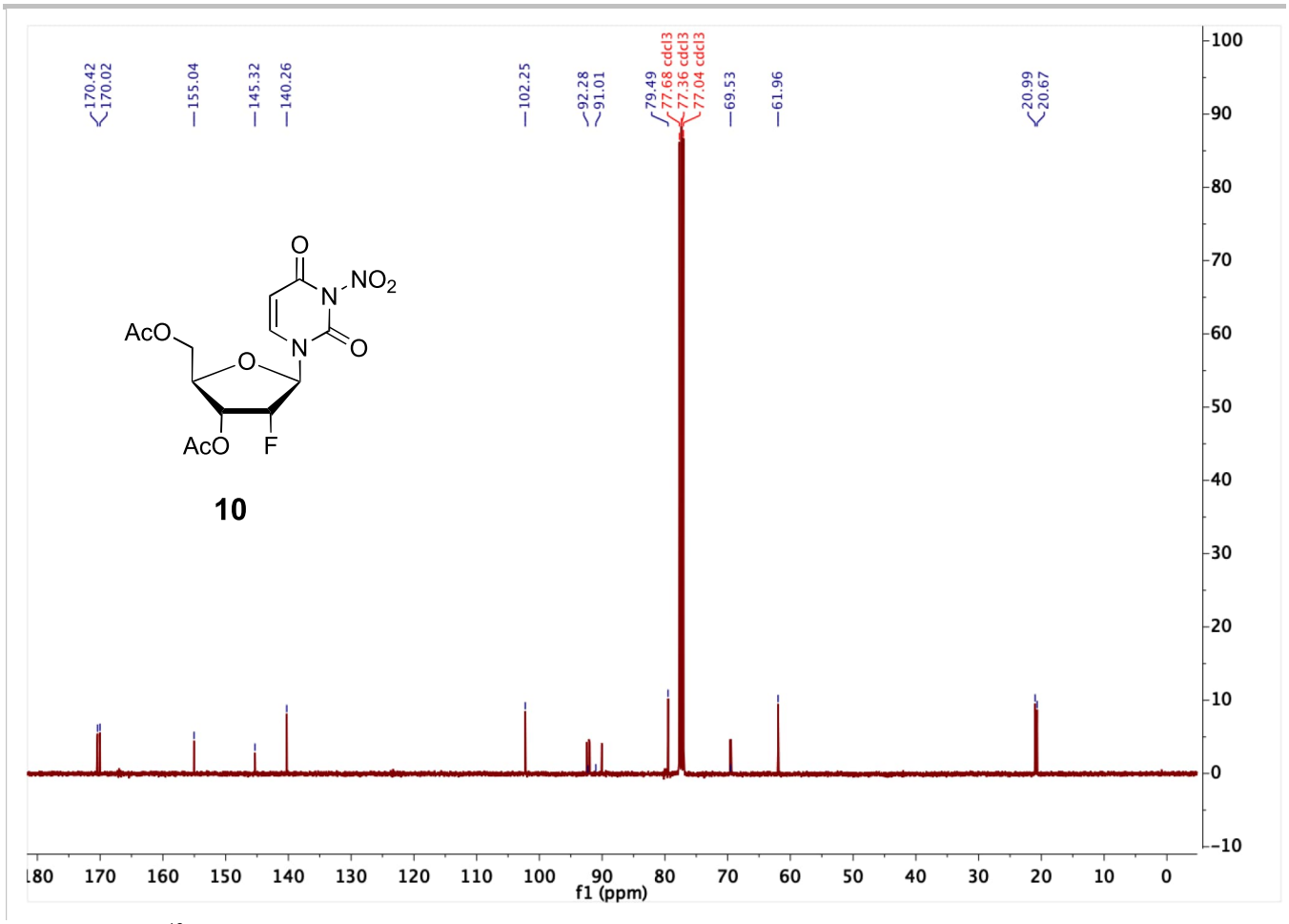


Figure SI-13: ^{13}C-NMR spectrum of compound **10**.

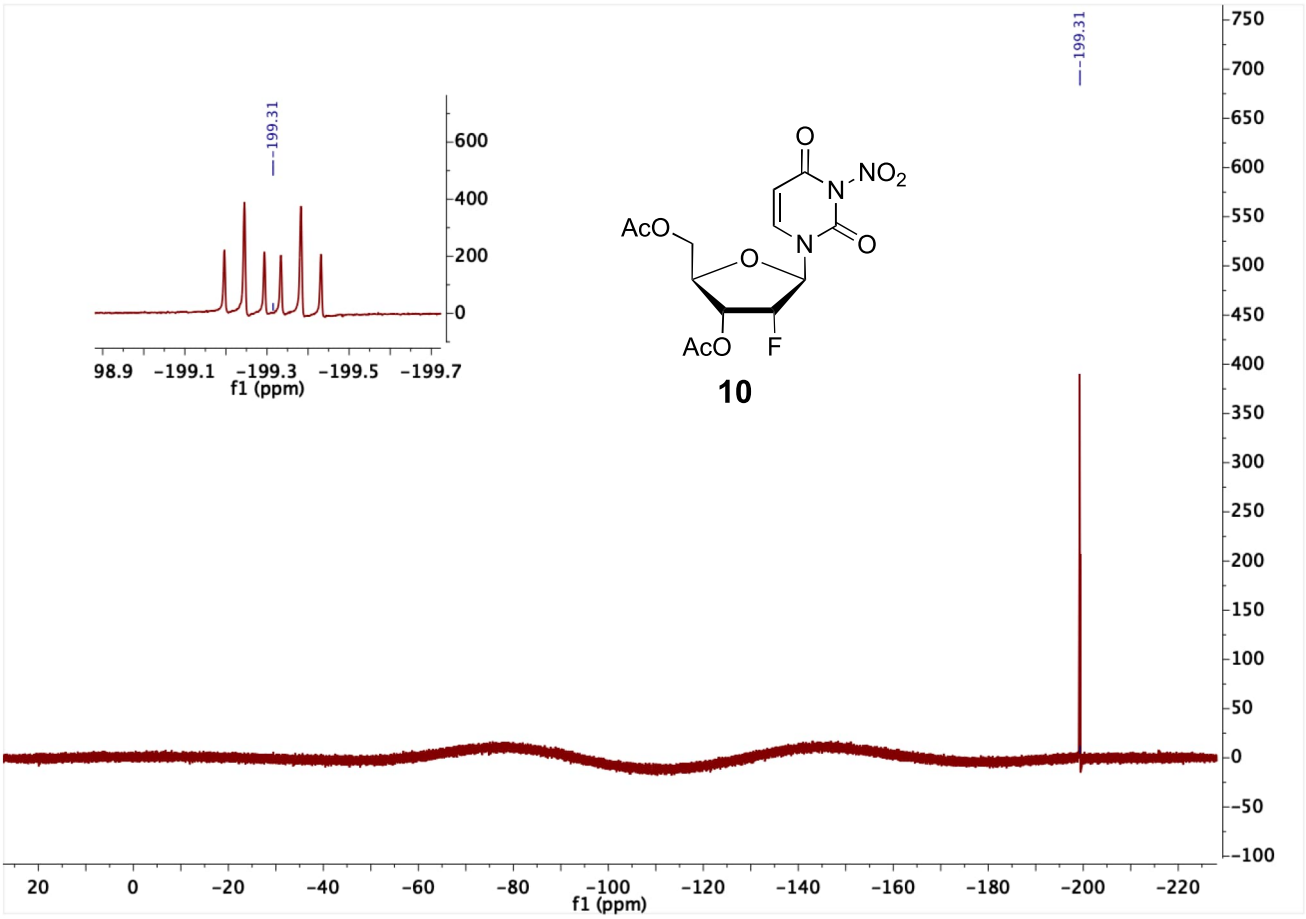


Figure SI-14: ^{19}F-NMR spectrum of compound **10**.

3',5'-di-*O*-Acetyl-2'-deoxy-2'-(*R*)-fluoro-(N^3-^{15}N)-uridine (11)

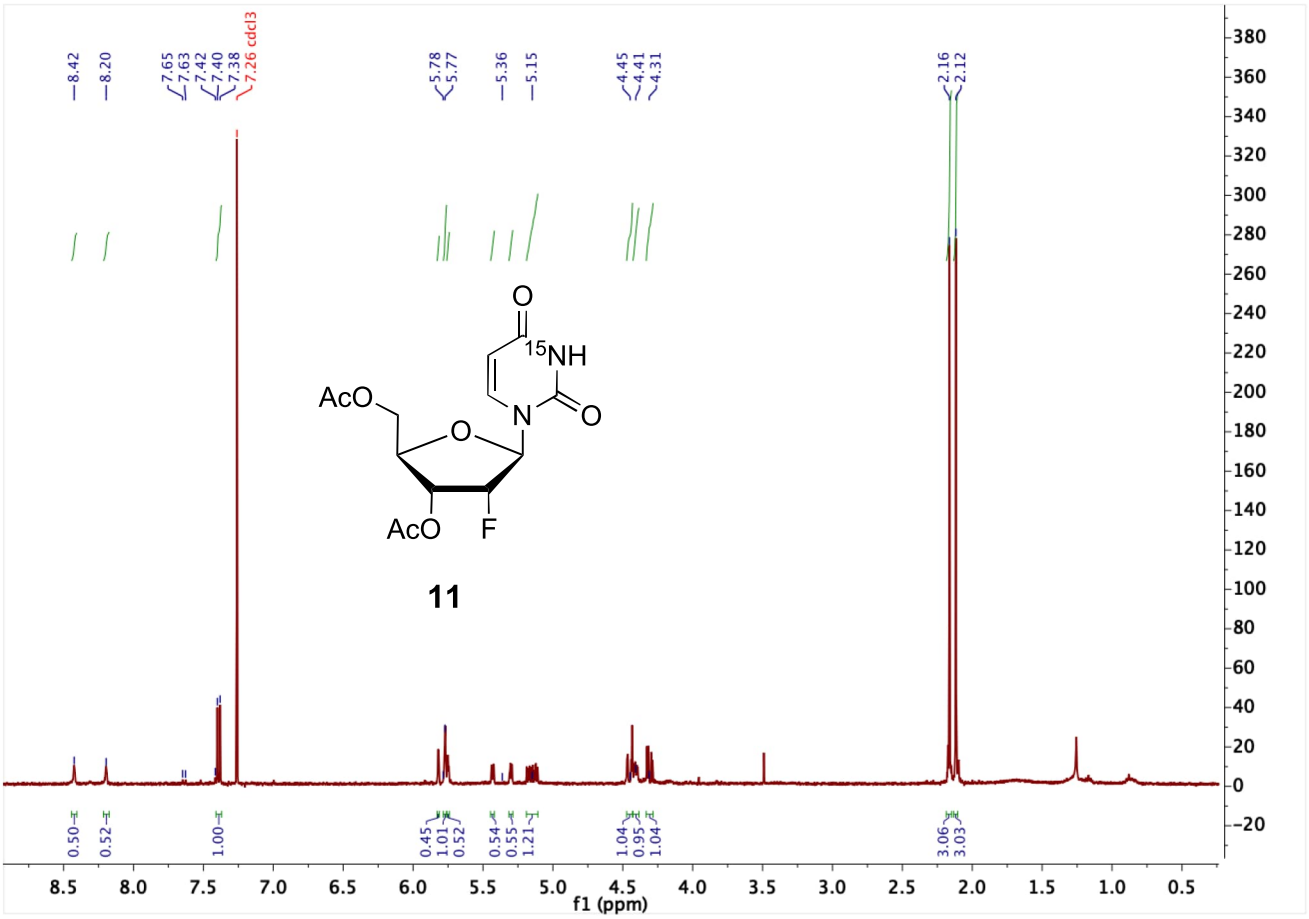


Figure SI-15: ^{1}H-NMR spectrum of compound **11**.

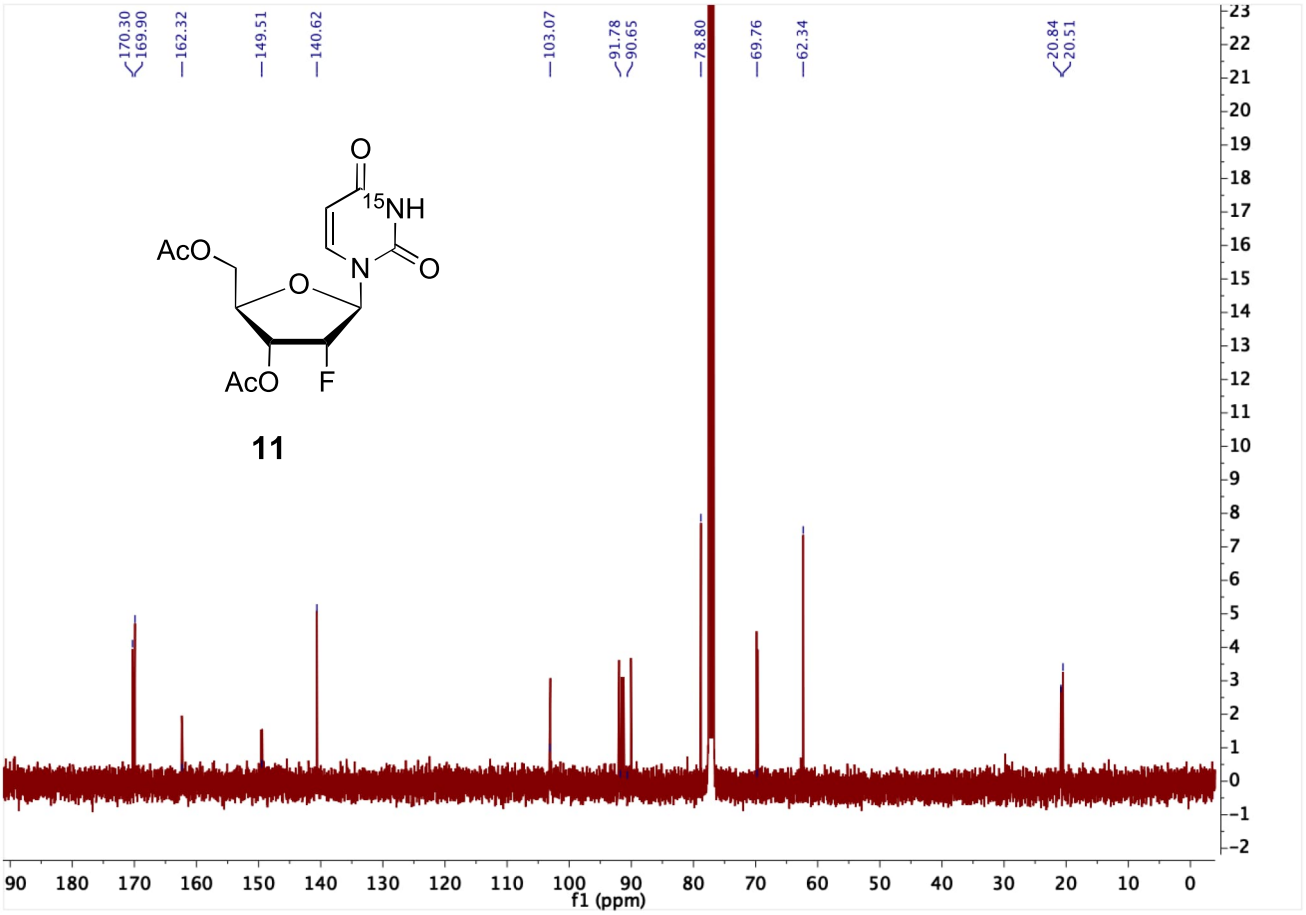


Figure SI-16: ^{13}C-NMR spectrum of compound **11**.

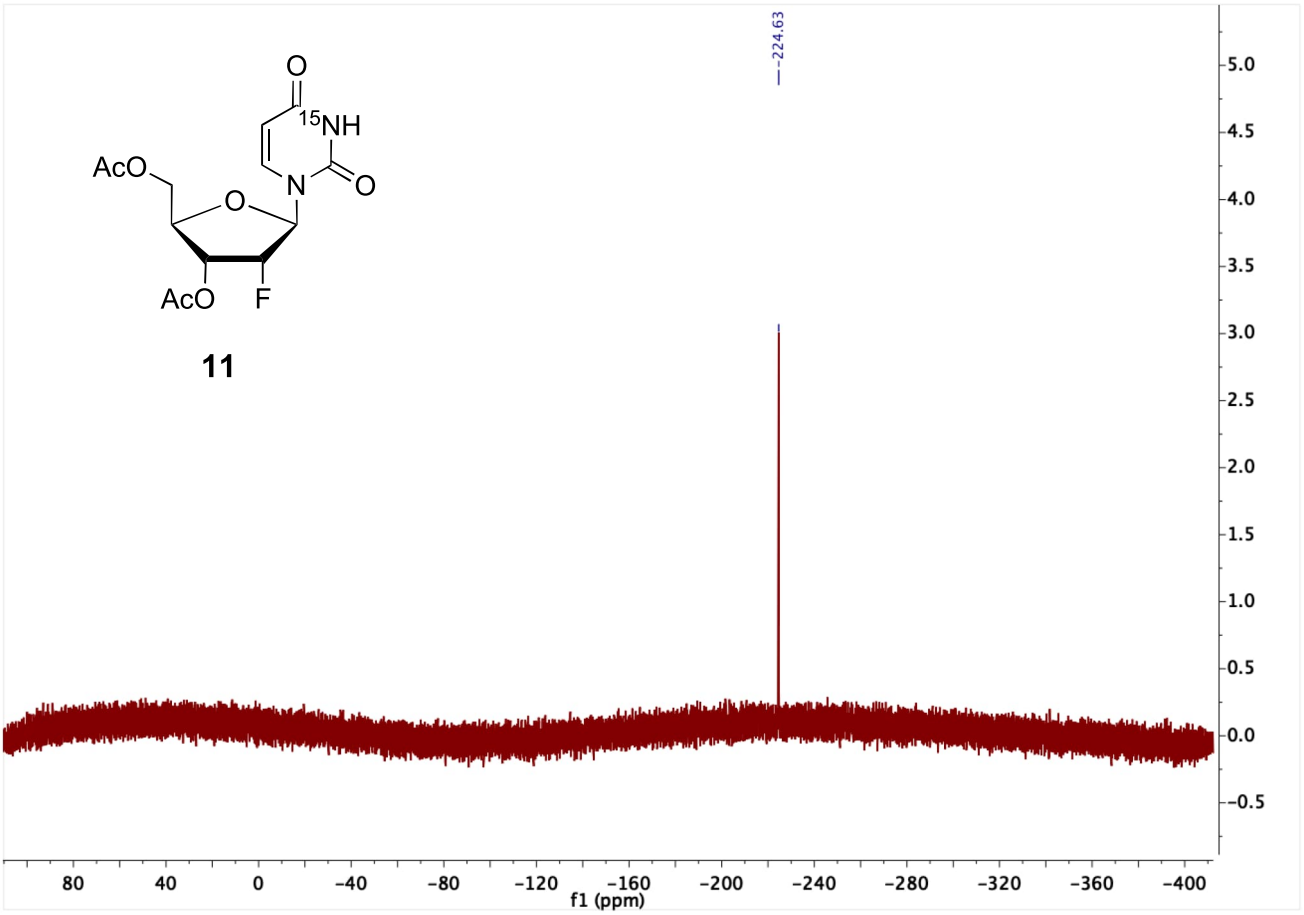


Figure SI-17: ^{15}N-NMR spectrum of compound **11**.

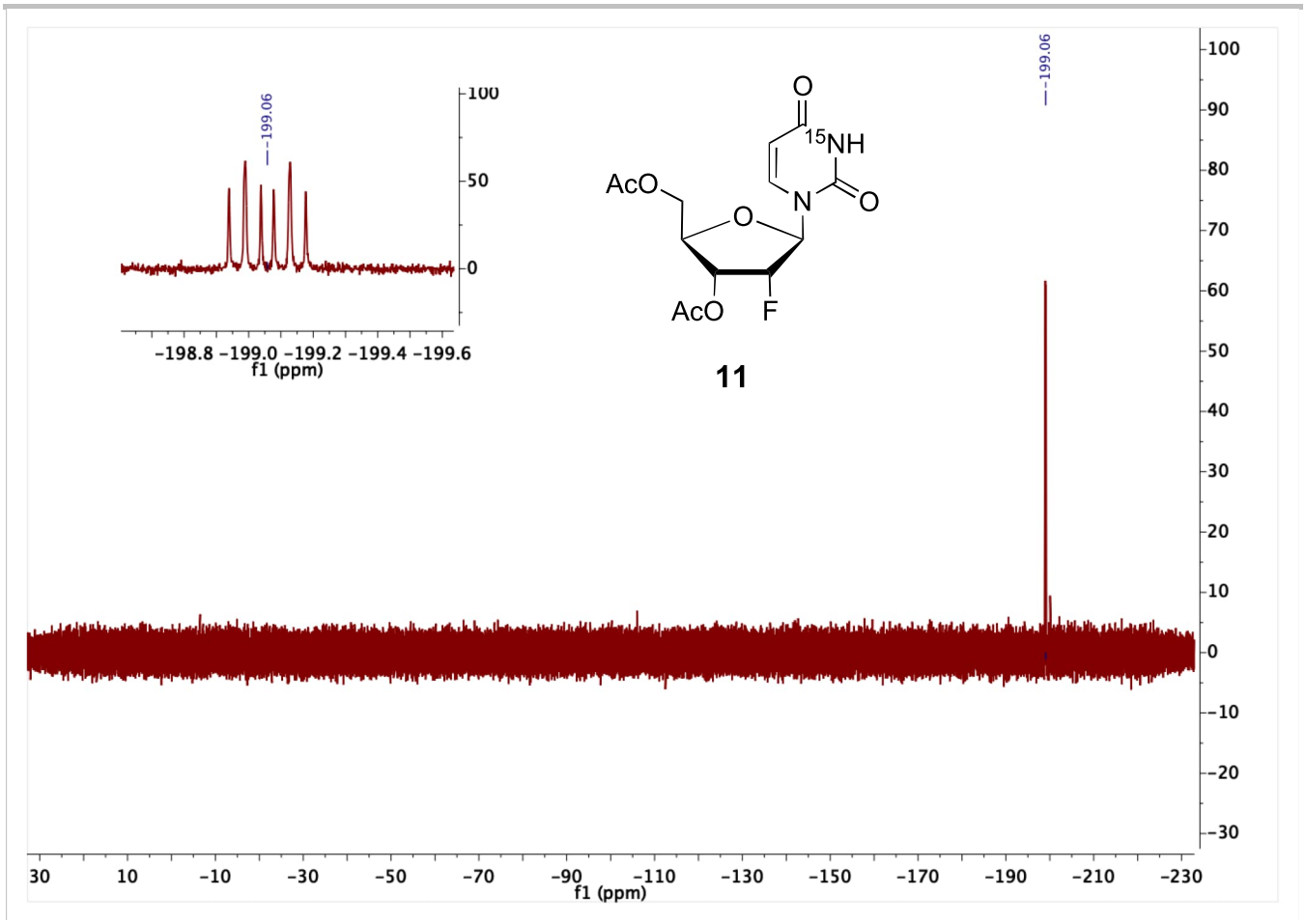


Figure SI-18: ^{19}F-NMR spectrum of compound **11**.

3',5'-bis-*O*-(*tert*-Butyl(dimethyl)silyl)-2'-deoxy-2'-(*R*)-fluoro-(N^3-^{15}N)-uridine (12)

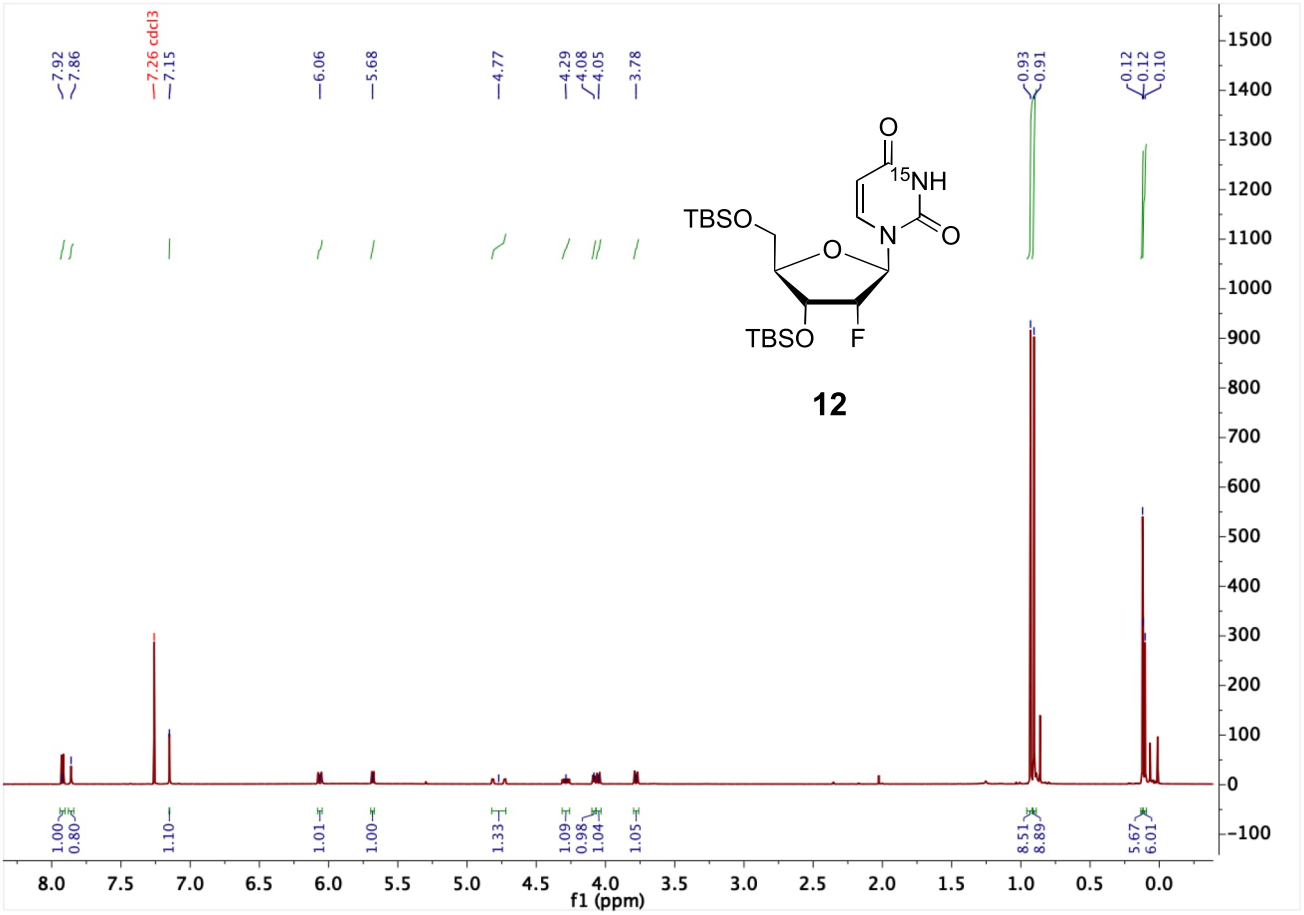


Figure SI-19: ^{1}H-NMR spectrum of compound **12**.

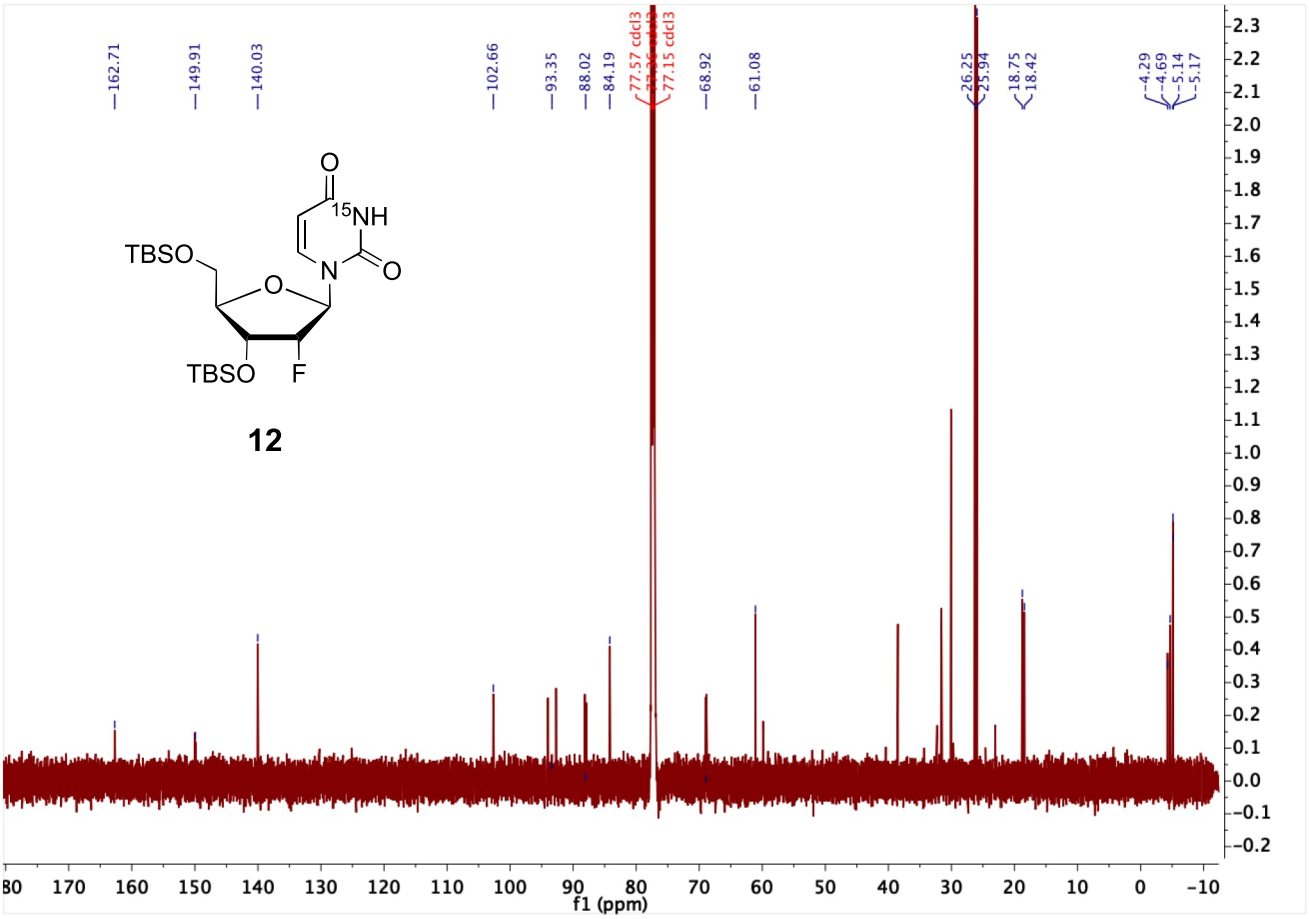


Figure SI-20: ^{13}C-NMR spectrum of compound **12**.

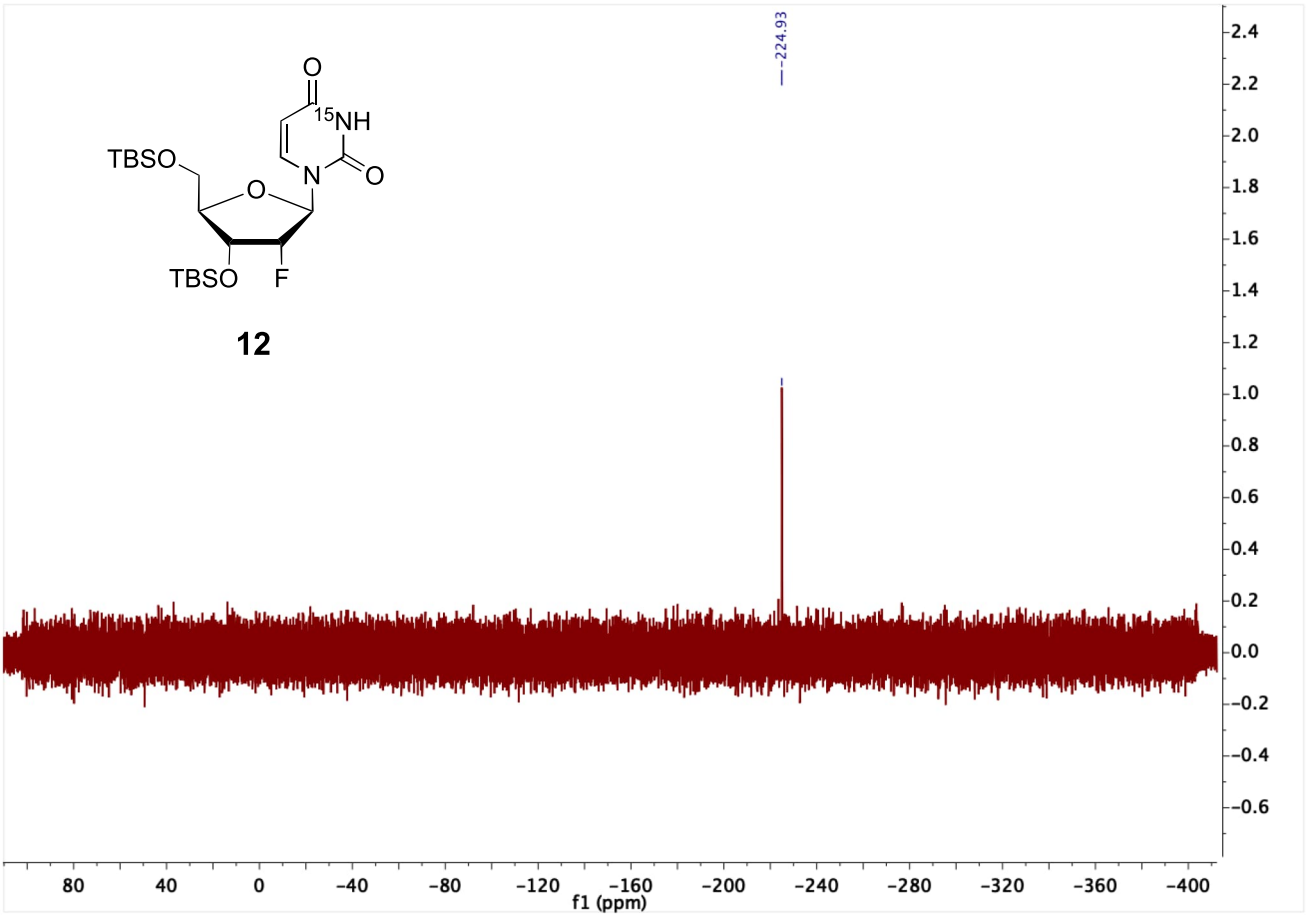


Figure SI-21: ^{15}N-NMR spectrum of compound **12**.

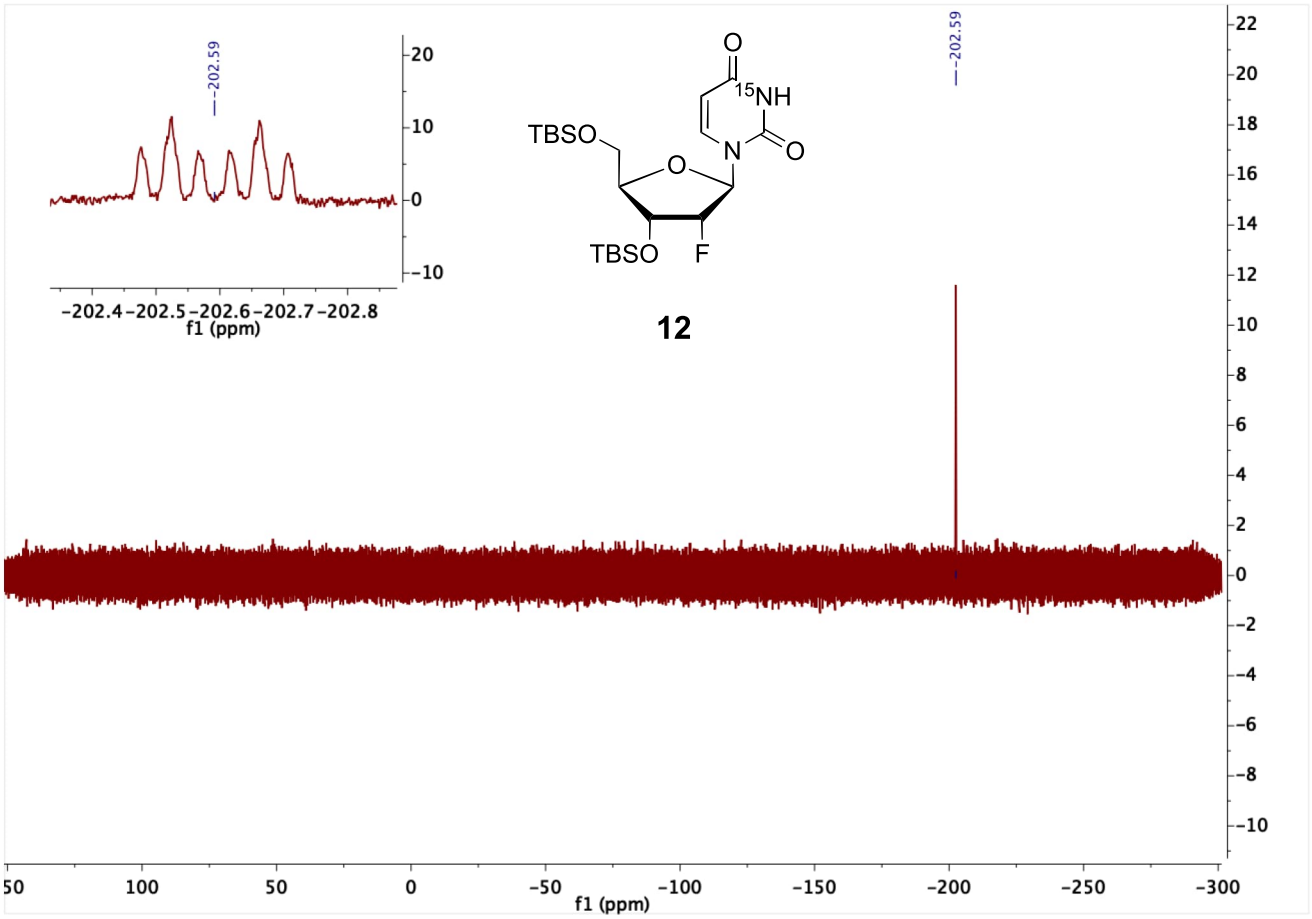


Figure SI-22: ^{19}F-NMR spectrum of compound **12**.

3',5'-bis-*O*-[*tert*-Butyl(dimethyl)silyl]-2'-deoxy-2'-(*R*)-fluoro-(N^3,N^4-$^{15}N_2$)-cytidine (13)

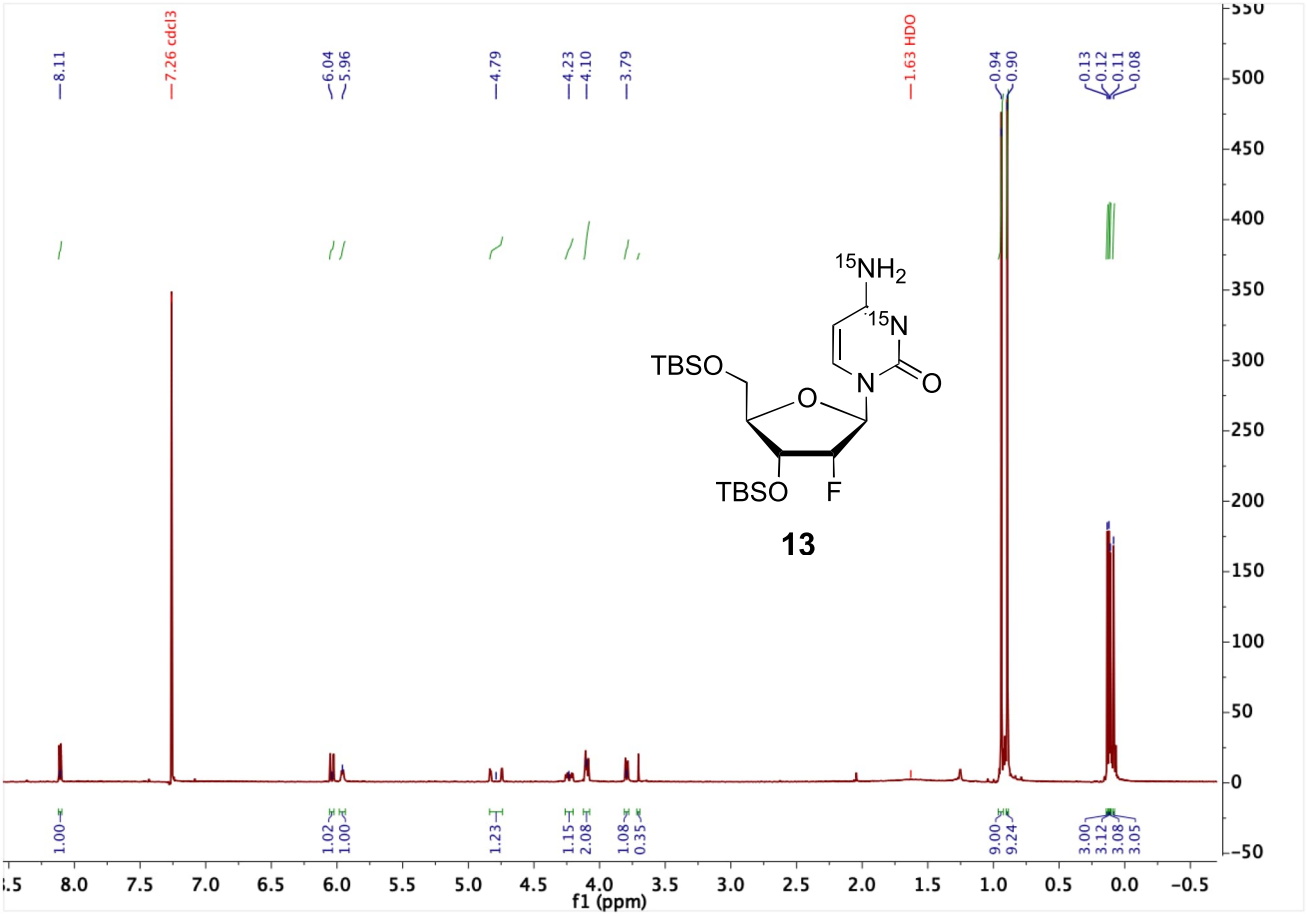


Figure SI-23: ^{1}H-NMR spectrum of compound **13**.

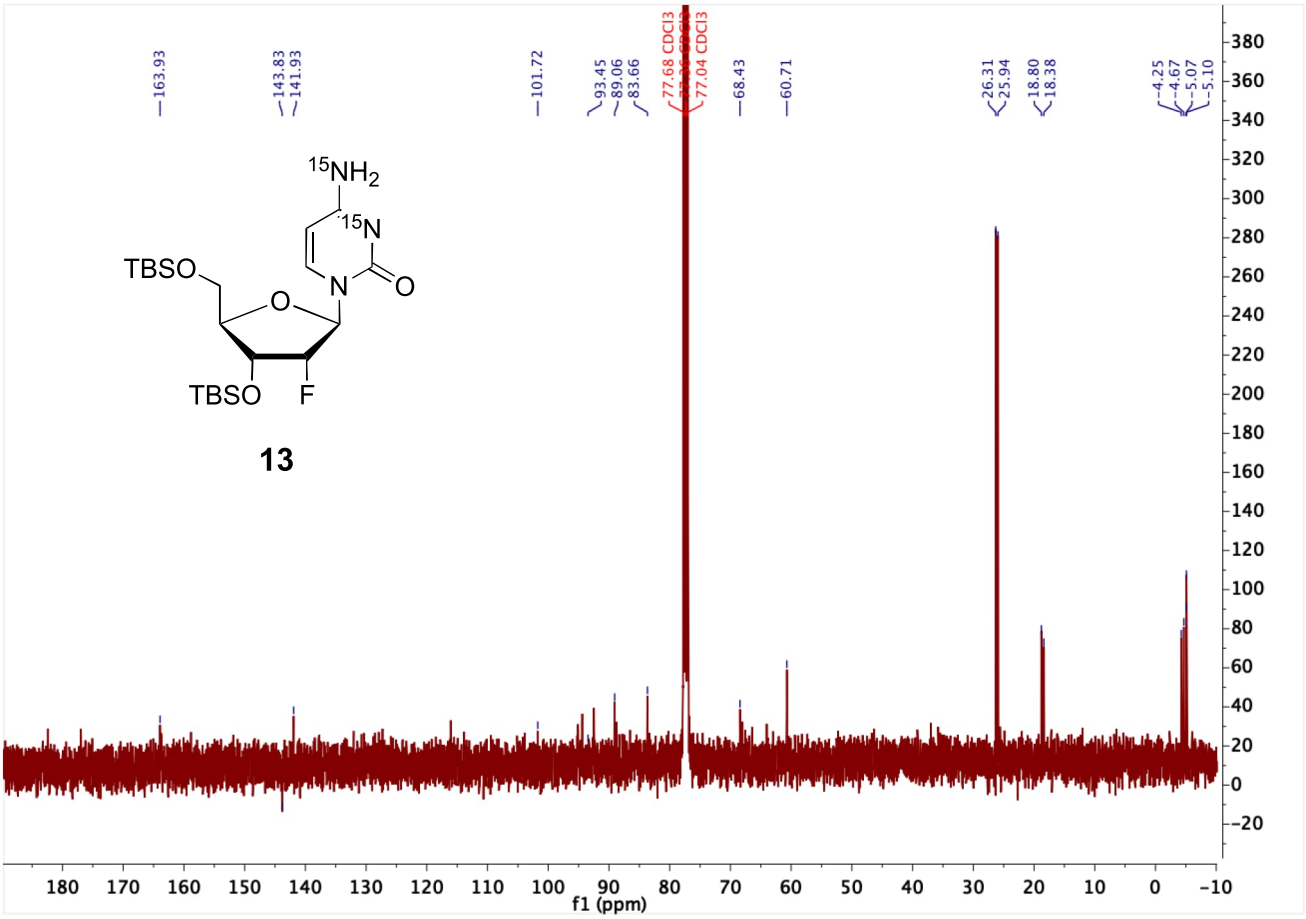


Figure SI-24: ^{13}C-NMR spectrum of compound **13**.

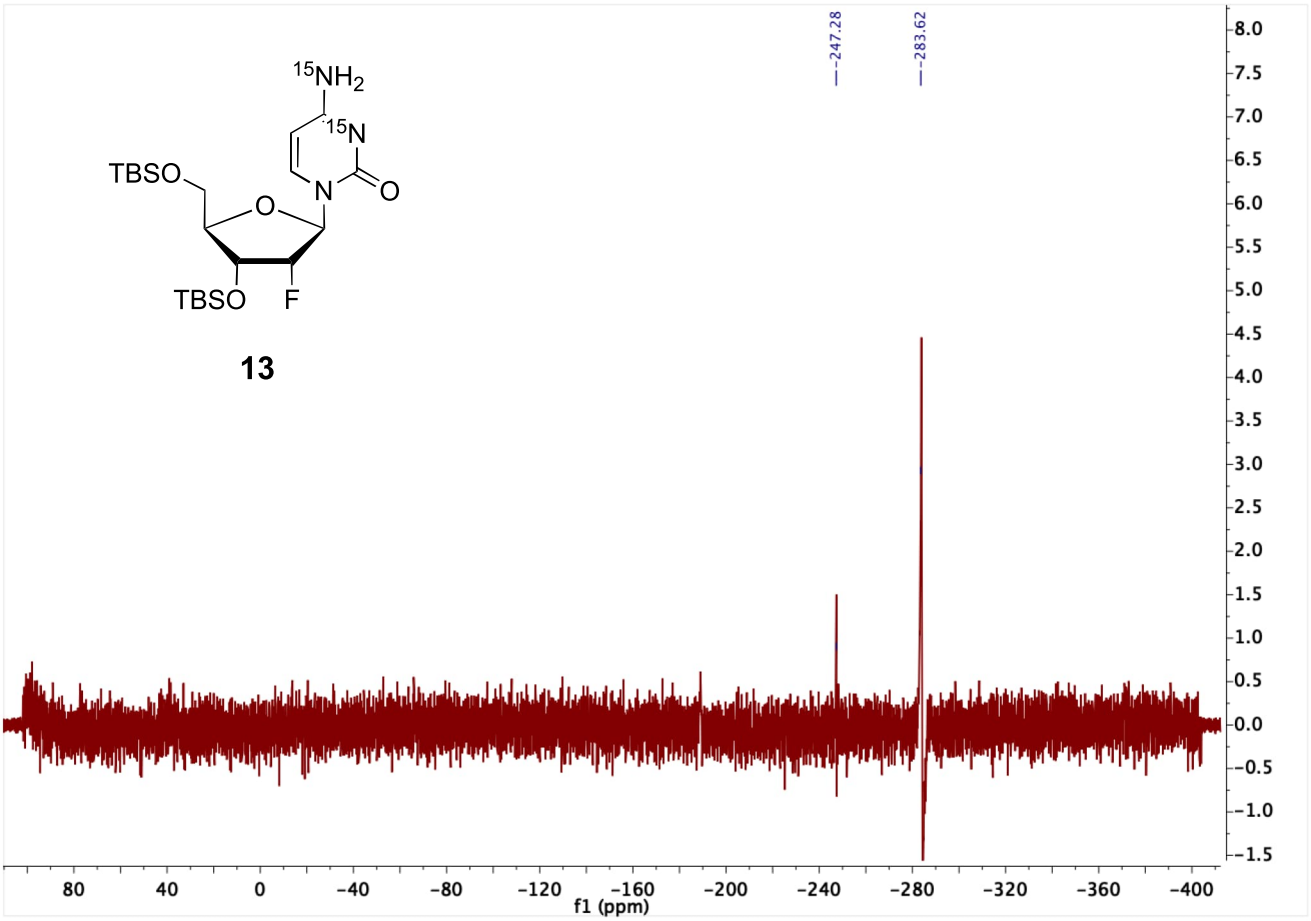


Figure SI-25: ^{15}N-NMR spectrum of compound **13**.

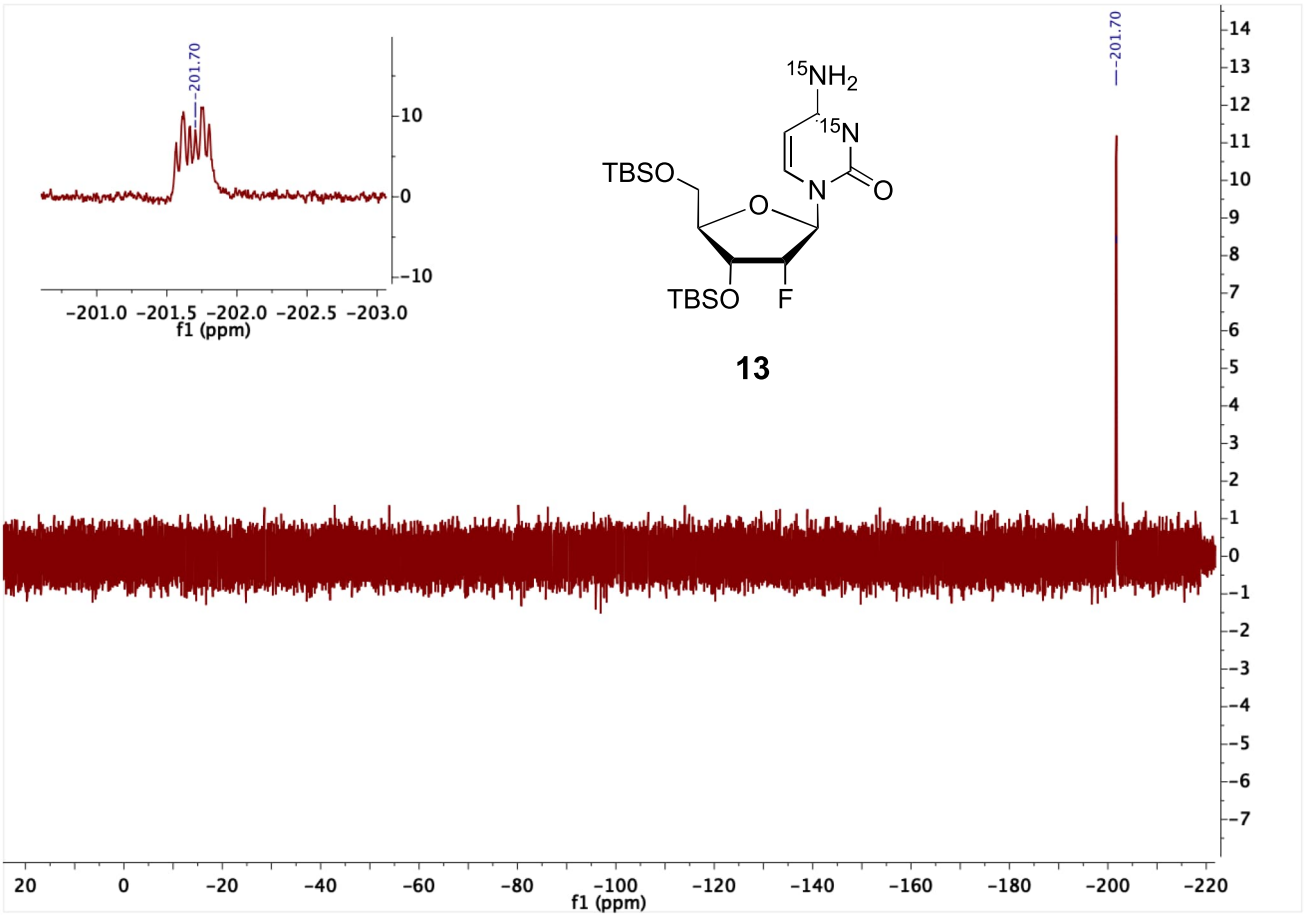


Figure SI-26: ^{19}F-NMR spectrum of compound **13**.

3',5'-bis-*O*-(*tert*-Butyl(dimethyl)silyl)-5-iodo-2'-deoxy-2'-(*R*)-fluoro-(N^3,N^4-$^{15}N_2$)-cytidine (14)

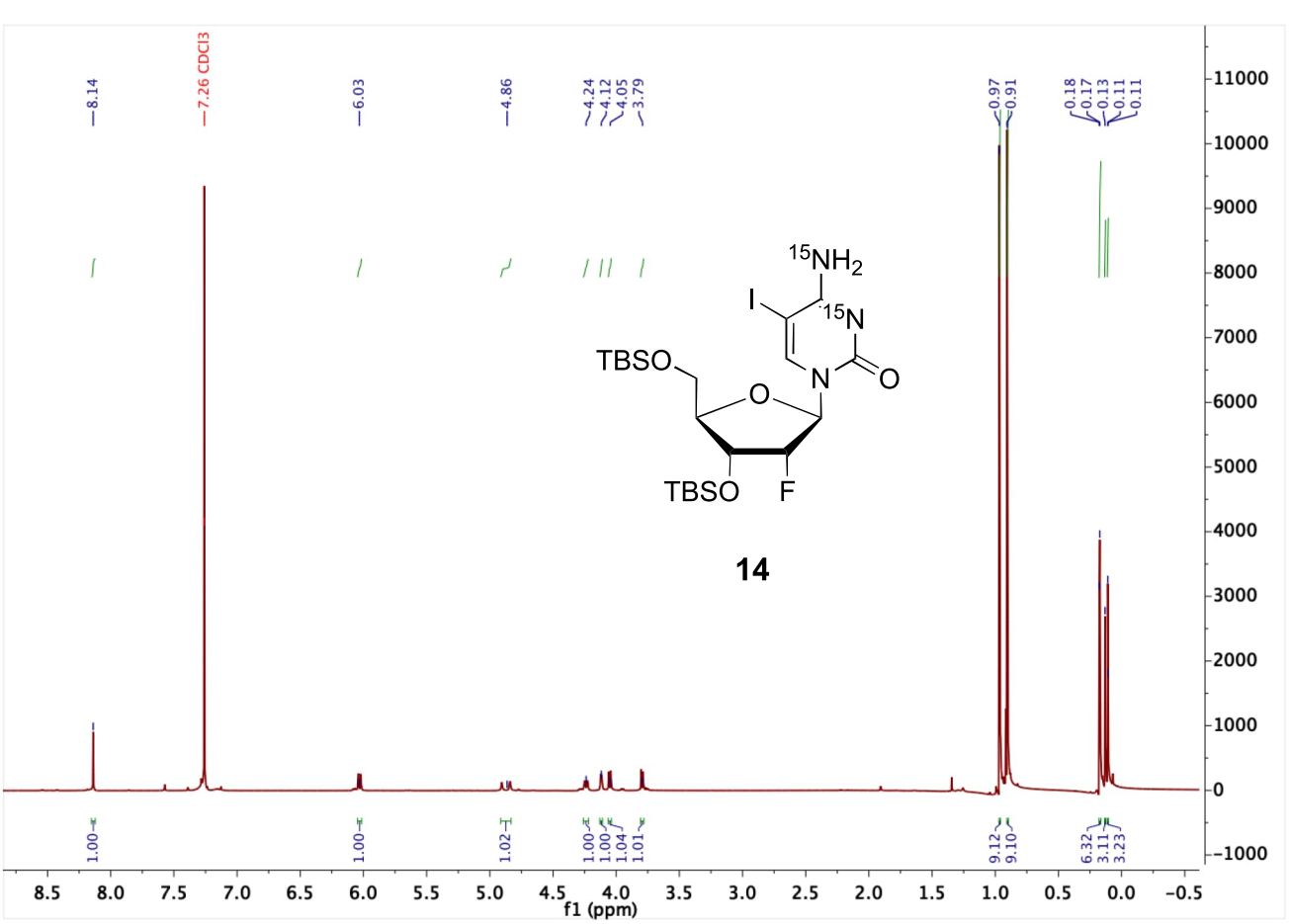


Figure SI-27: ^{1}H-NMR spectrum of compound **14**.

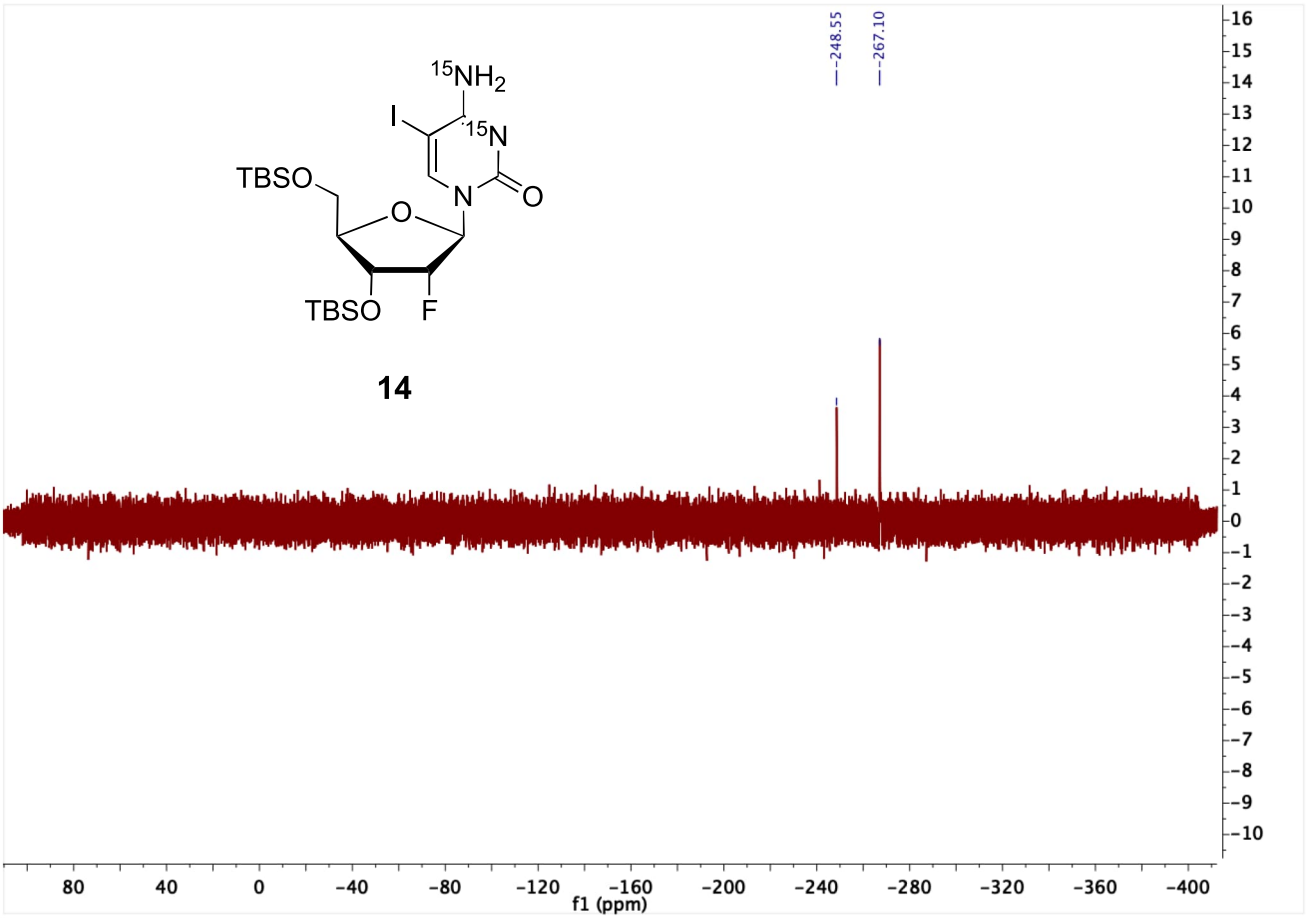


Figure SI-28: ^{15}N-NMR spectrum of compound **14**.

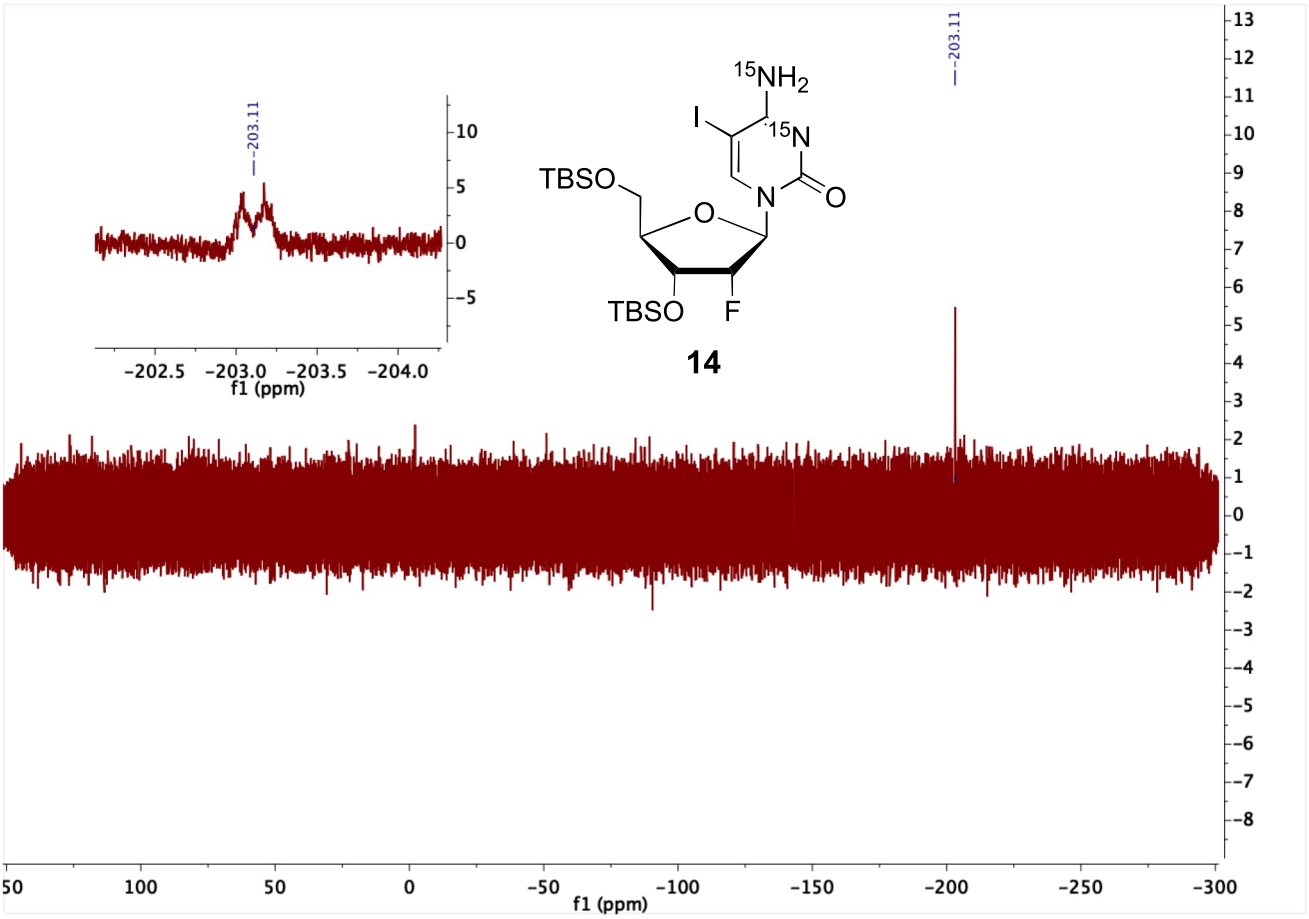


Figure SI-29: ^{19}F-NMR spectrum of compound **14**.

3',5'-bis-*O*-(*tert*-Butyl(dimethyl)silyl)-5-(methylcarboxy)-2'-deoxy-2'-(*R*)-fluoro-(N^3,N^4-$^{15}N_2$)-cytidine (15)

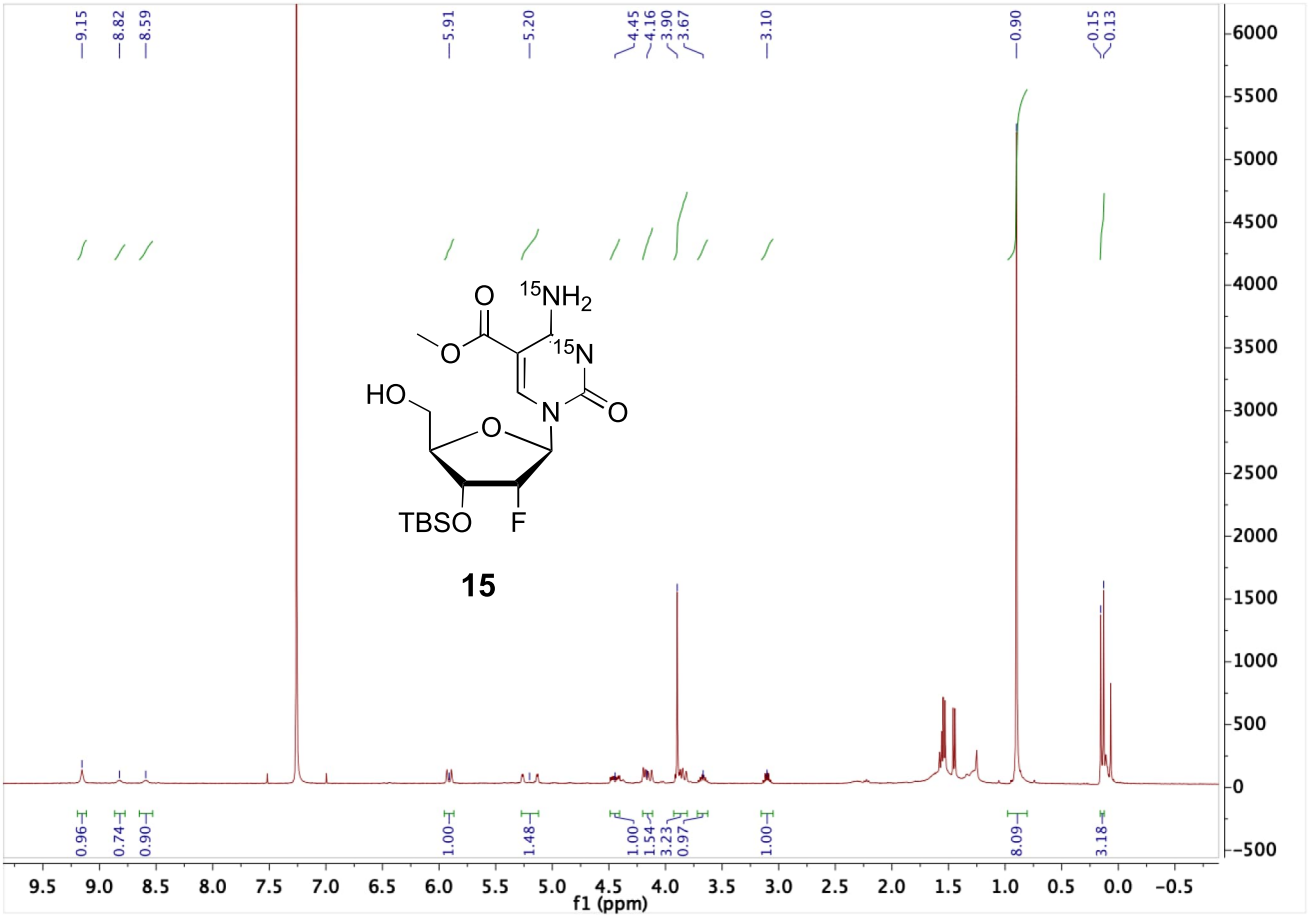


Figure SI-30: ^{1}H-NMR spectrum of compound **15**.

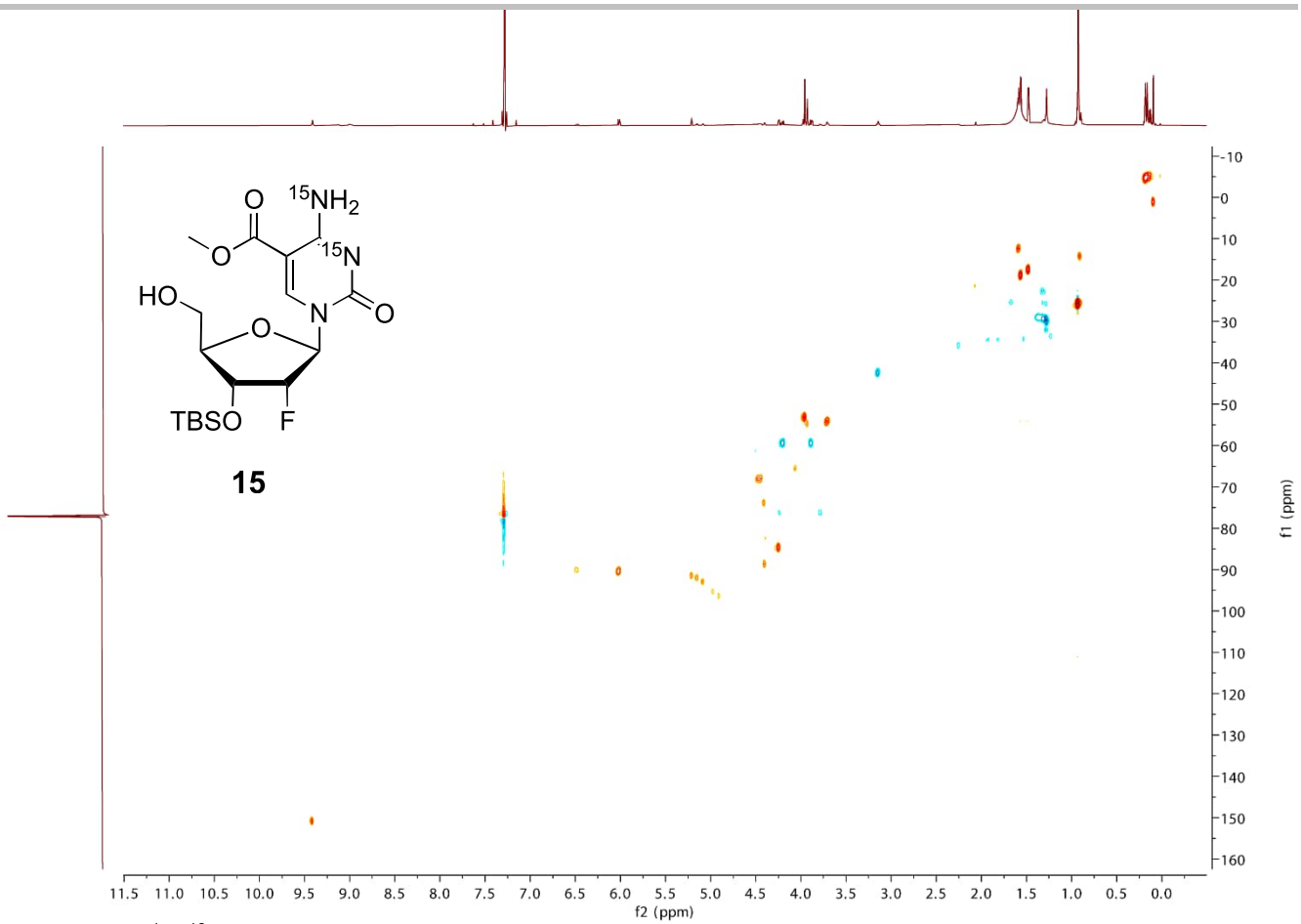


Figure SI-31: ^{1}H-^{13}C -HSQC-NMR spectrum of compound **15**.

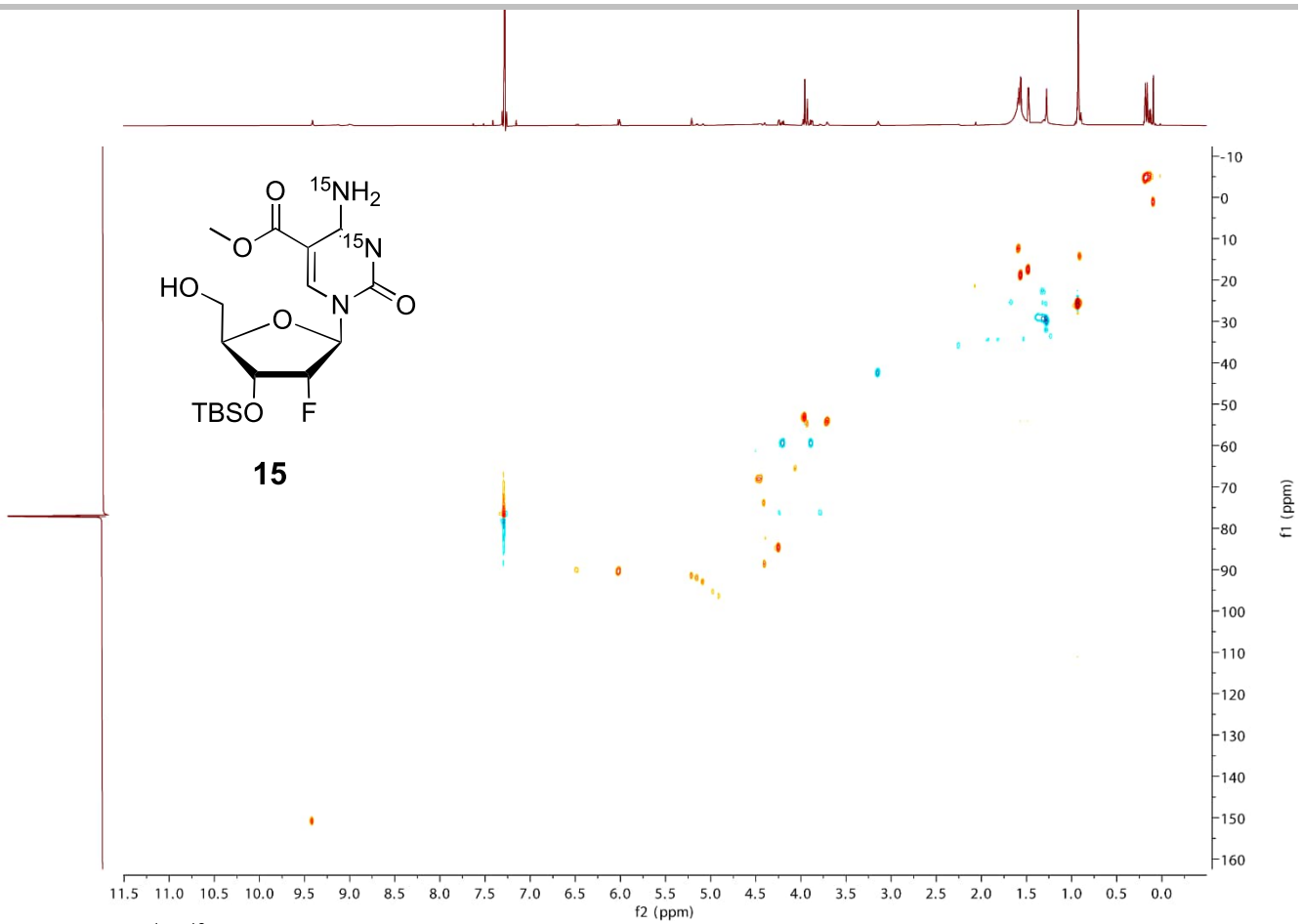


Figure SI-32: ^{1}H-^{13}C -HMBC-NMR spectrum of compound **15**.

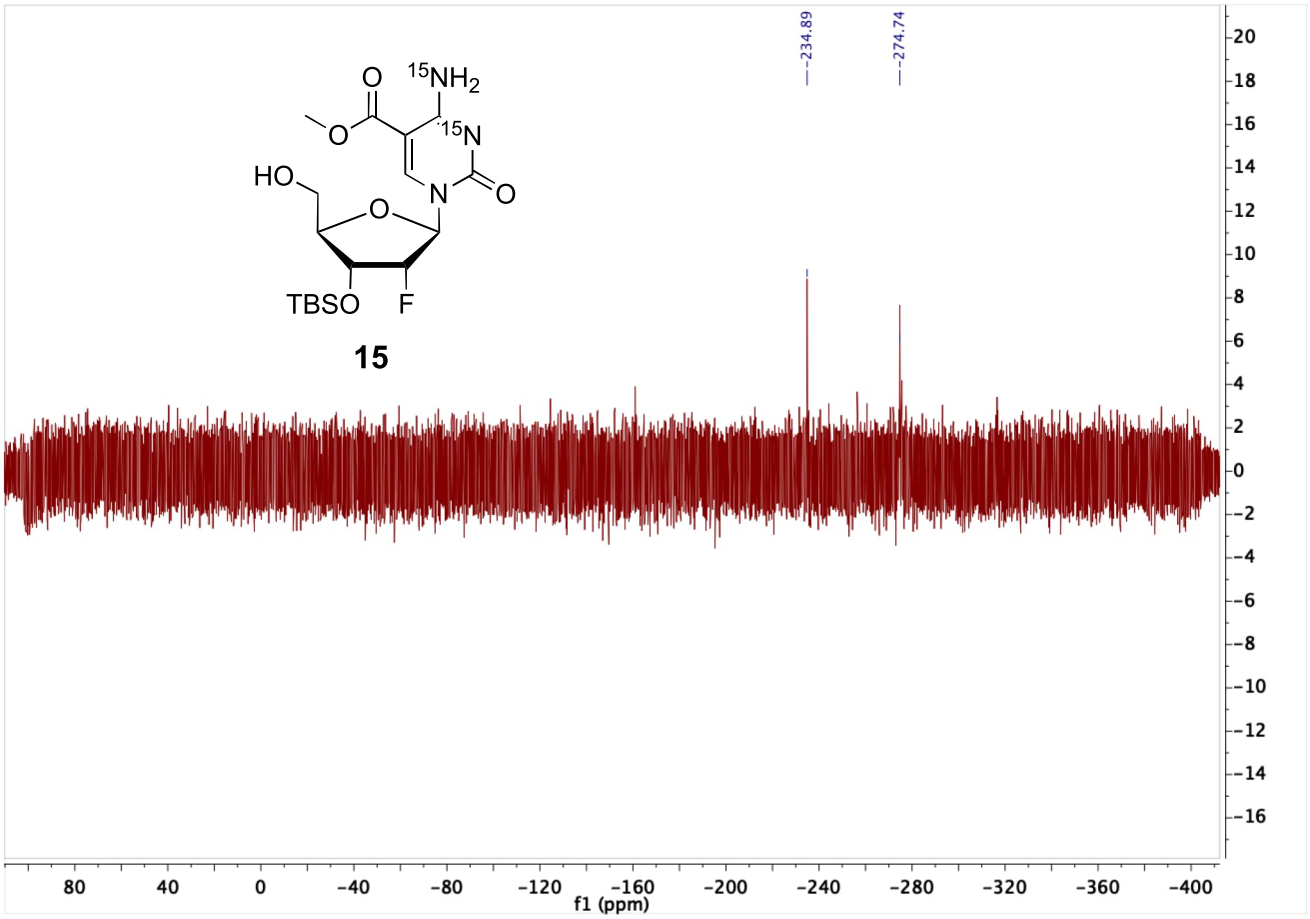


Figure SI-33: ^{15}N-NMR spectrum of compound **15**.

5-(Carboxy)-2'-deoxy-2'-(*R*)-fluoro-(N^3,N^4-$^{15}N_2$)-cytidine (3)

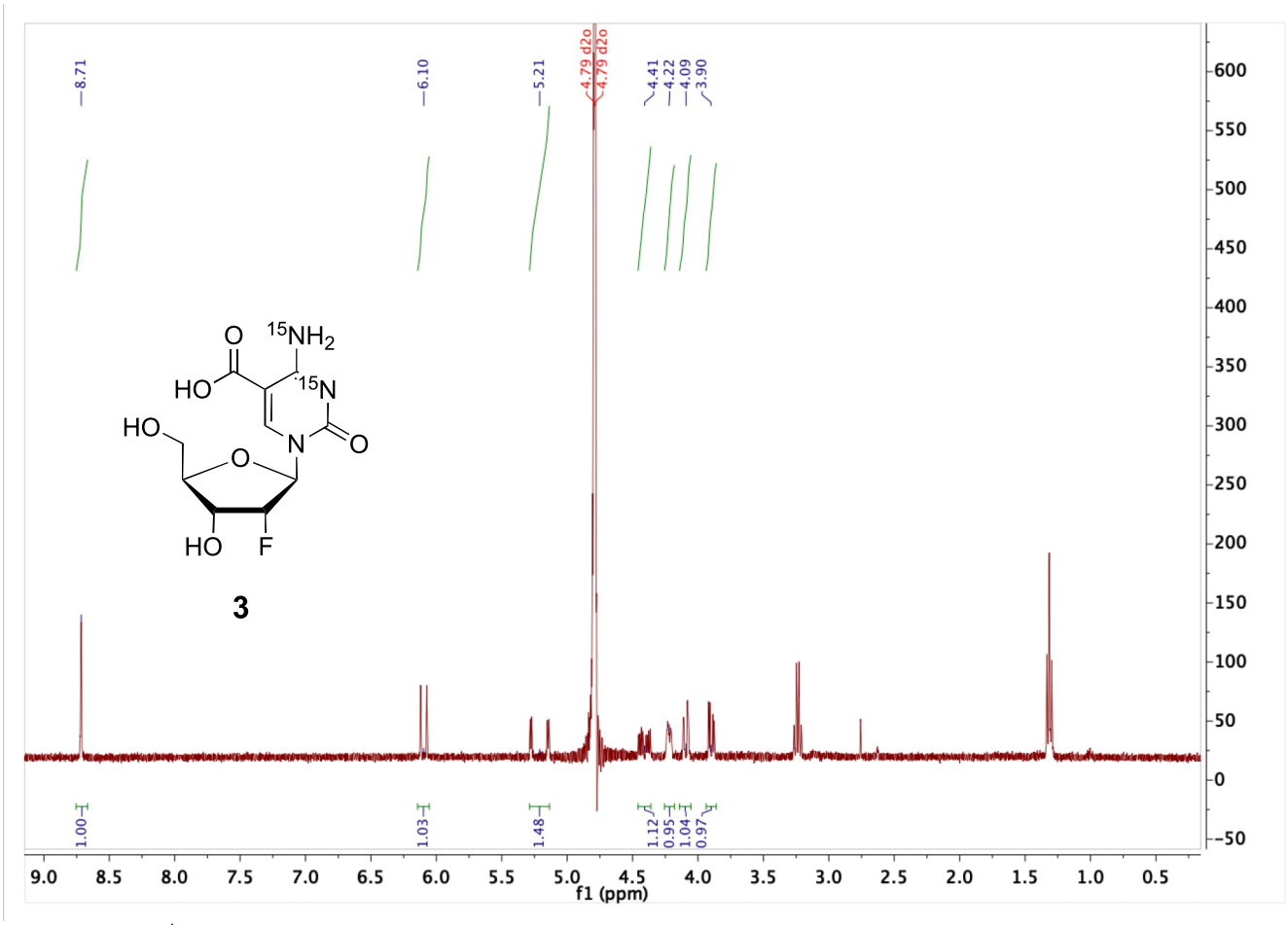


Figure SI-34: ^{1}H-NMR spectrum of compound **3**.

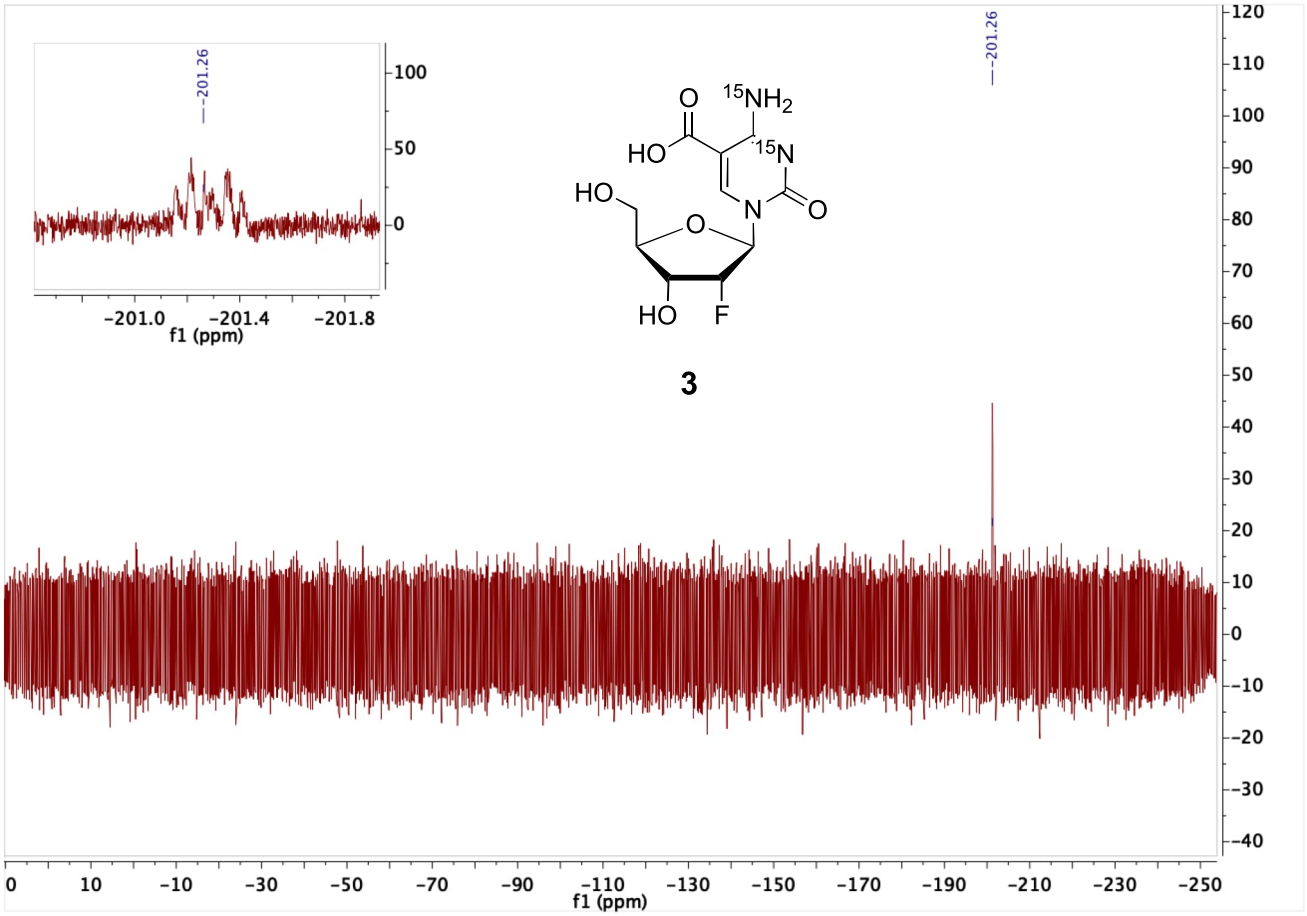


Figure SI-35: ^{19}F-NMR spectrum of compound **3**.

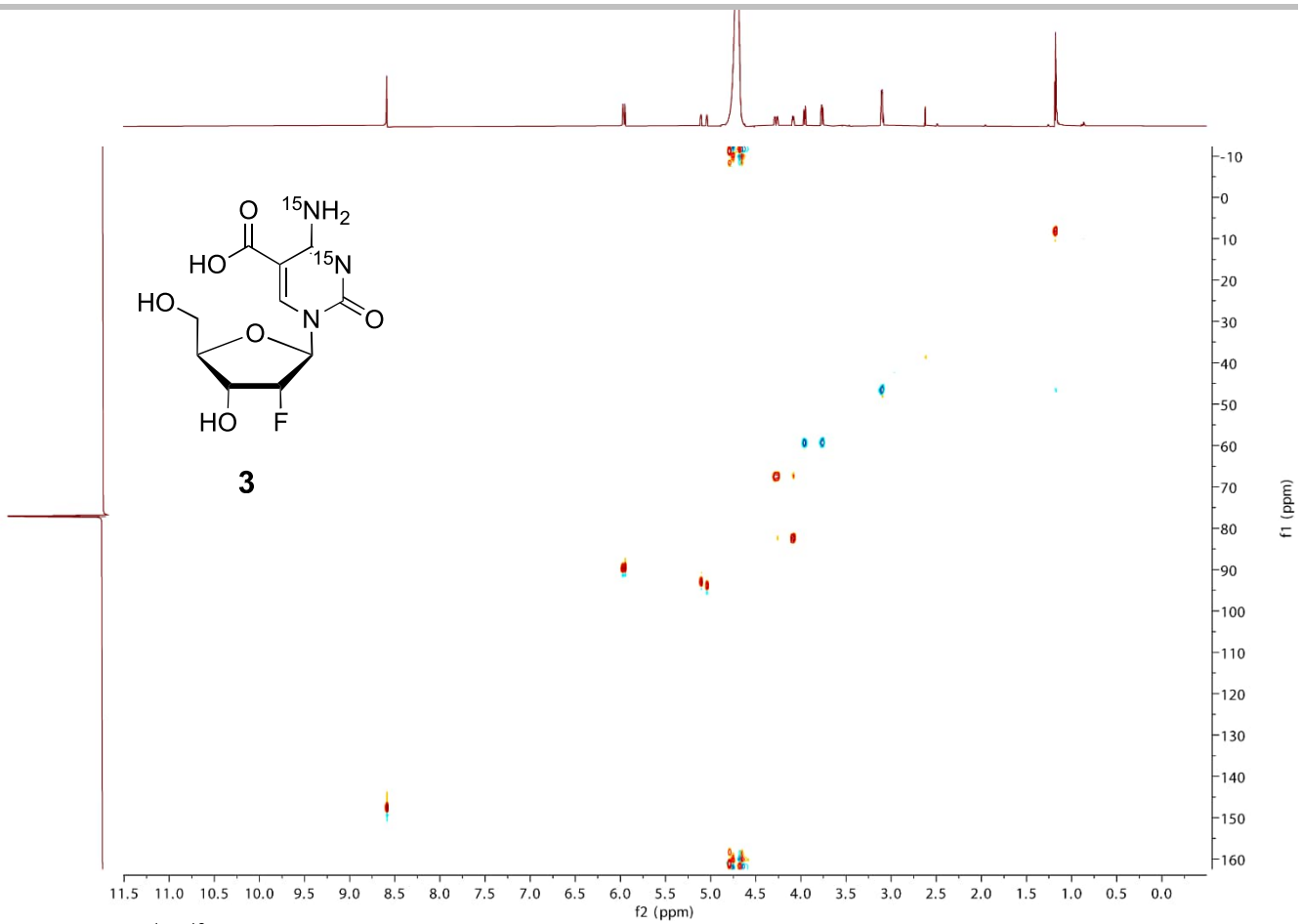


Figure SI-36: ^{1}H-^{13}C -HSQC-NMR spectrum of compound **3**.

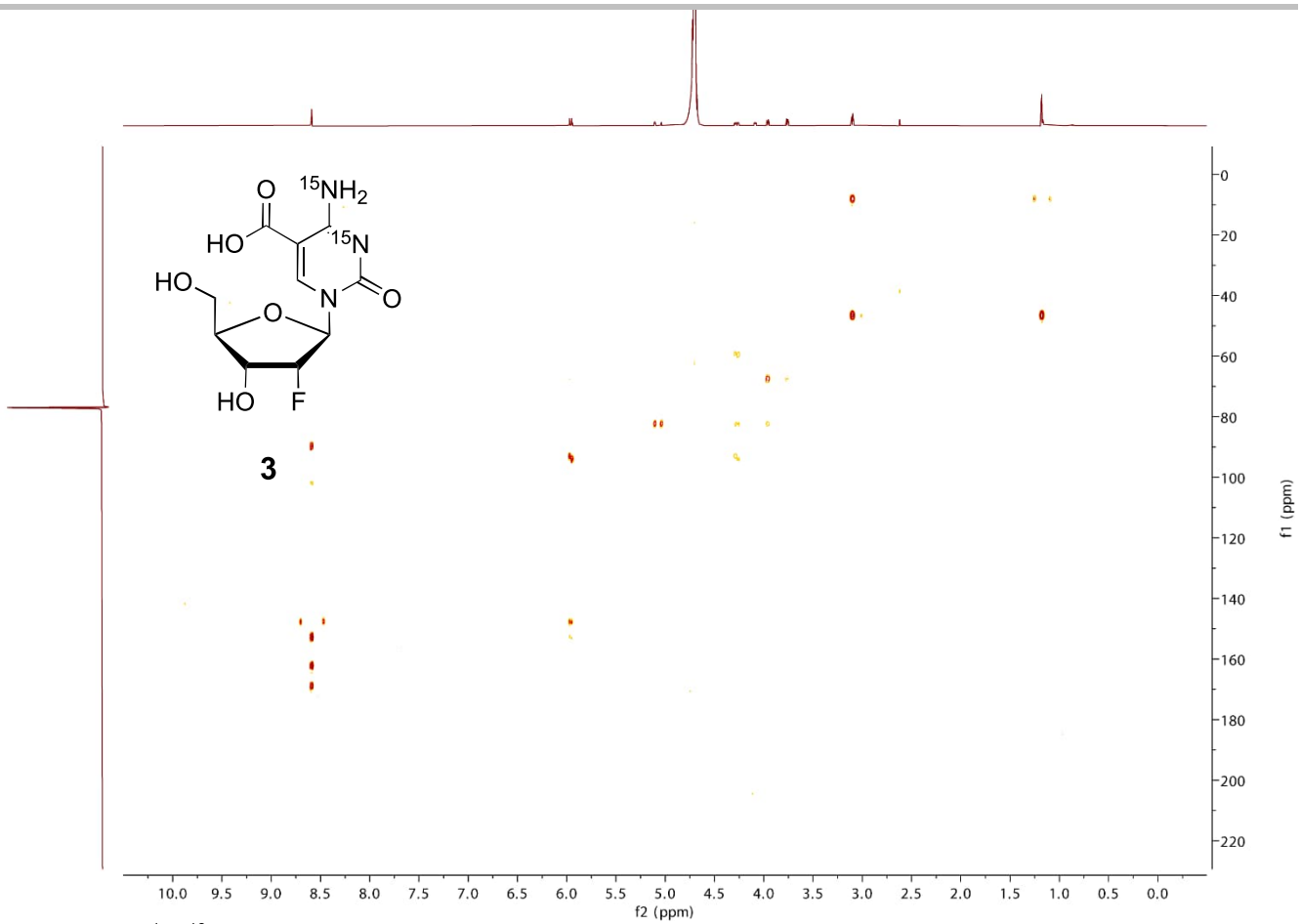


Figure SI-37: ^{1}H-^{13}C -HMBC-NMR spectrum of compound **3**.

3'-5'-*O*-bis(*tert*-Butyldimethylsilyl)-2'-desoxy-2'-(R)-fluoro-cytidine (16)

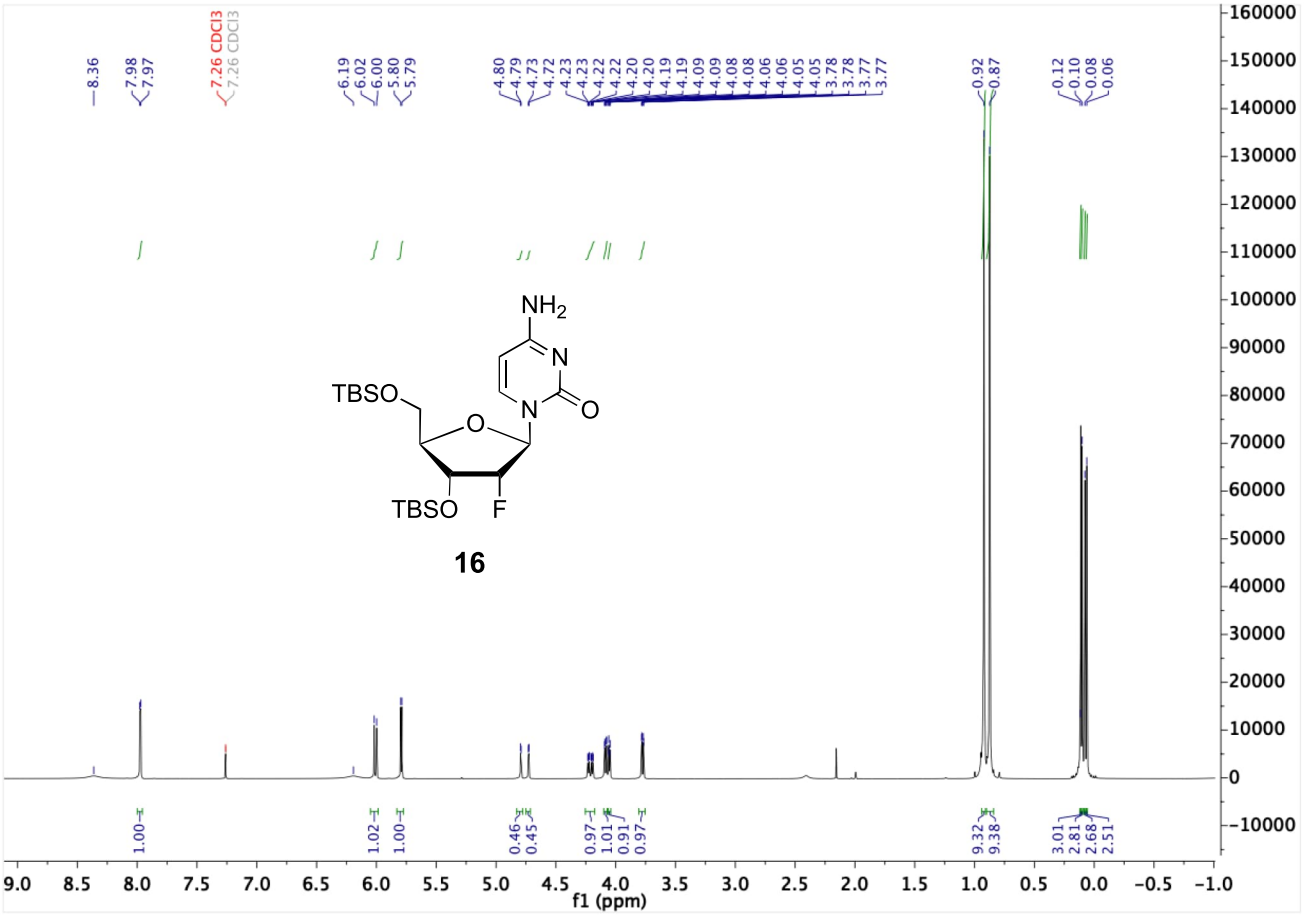


Figure SI-38: ^{1}H-NMR spectrum of compound **16**.

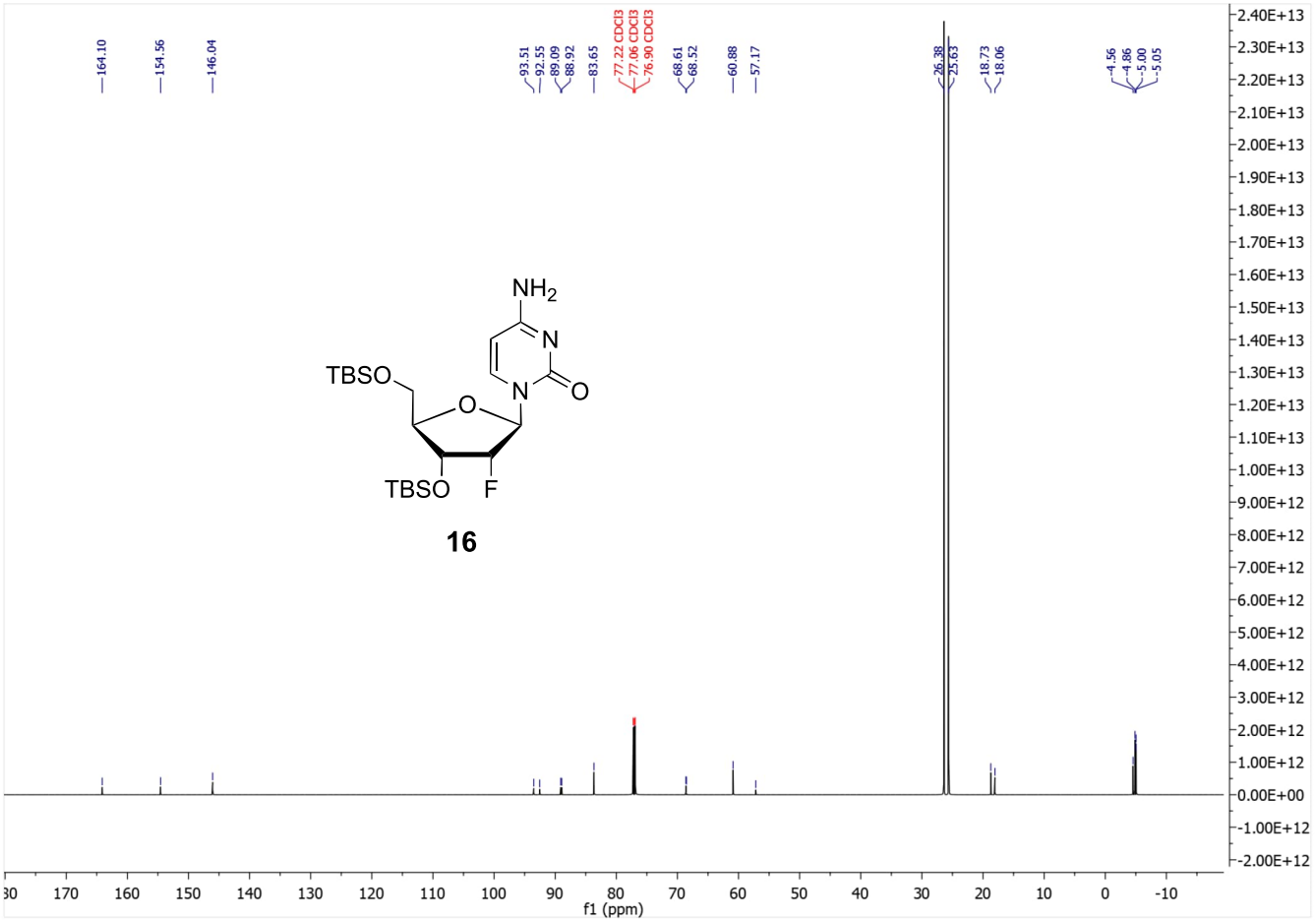


Figure SI-39: ^{13}C-NMR spectrum of compound **16**.

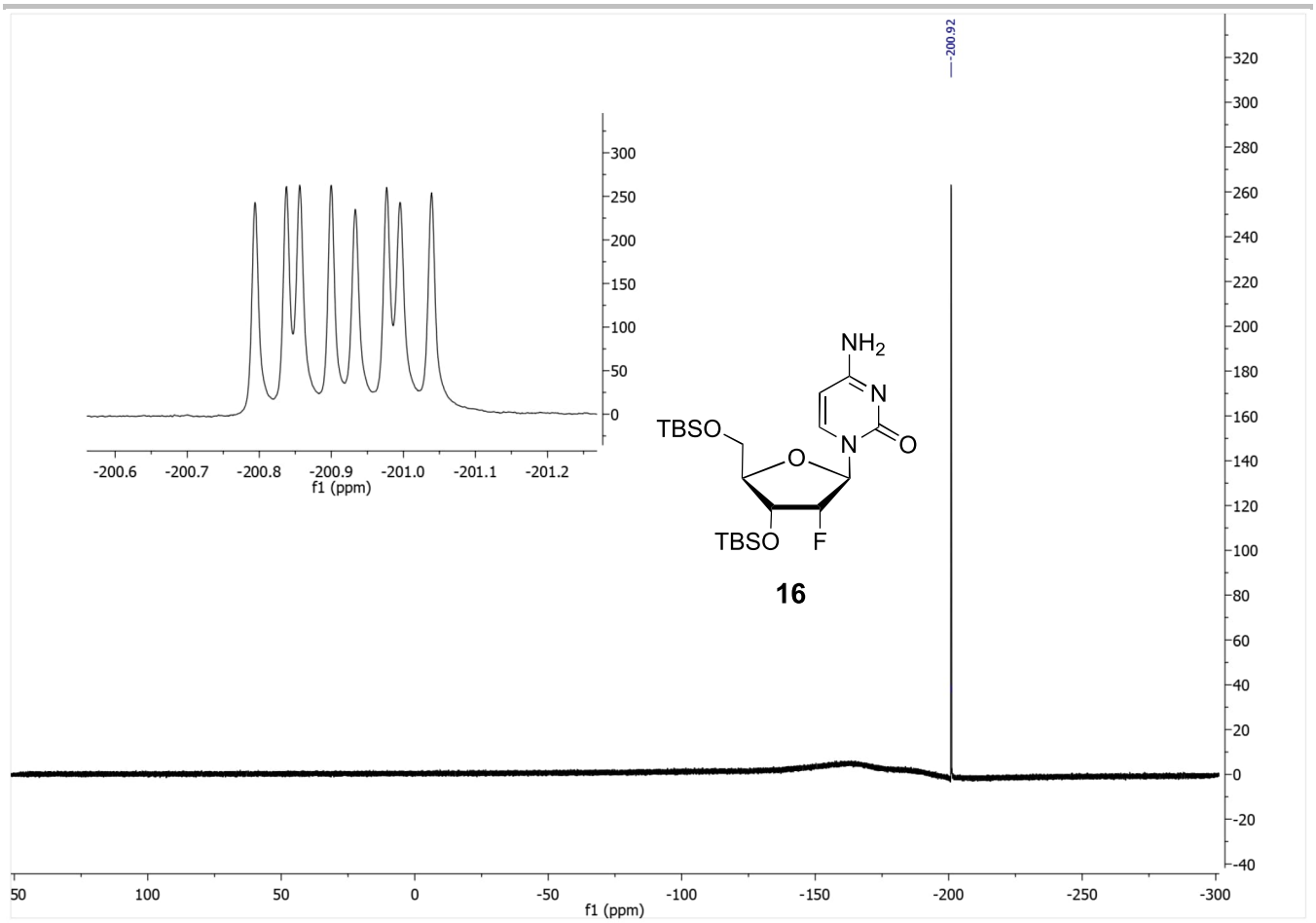


Figure SI-40: ^{19}F-NMR spectrum of compound **16**.

3'-5'-*O*-bis(*tert*-Butyldimethylsilyl)-5-iodo-2'-desoxy-2'-(R)-fluoro-cytidine (17)

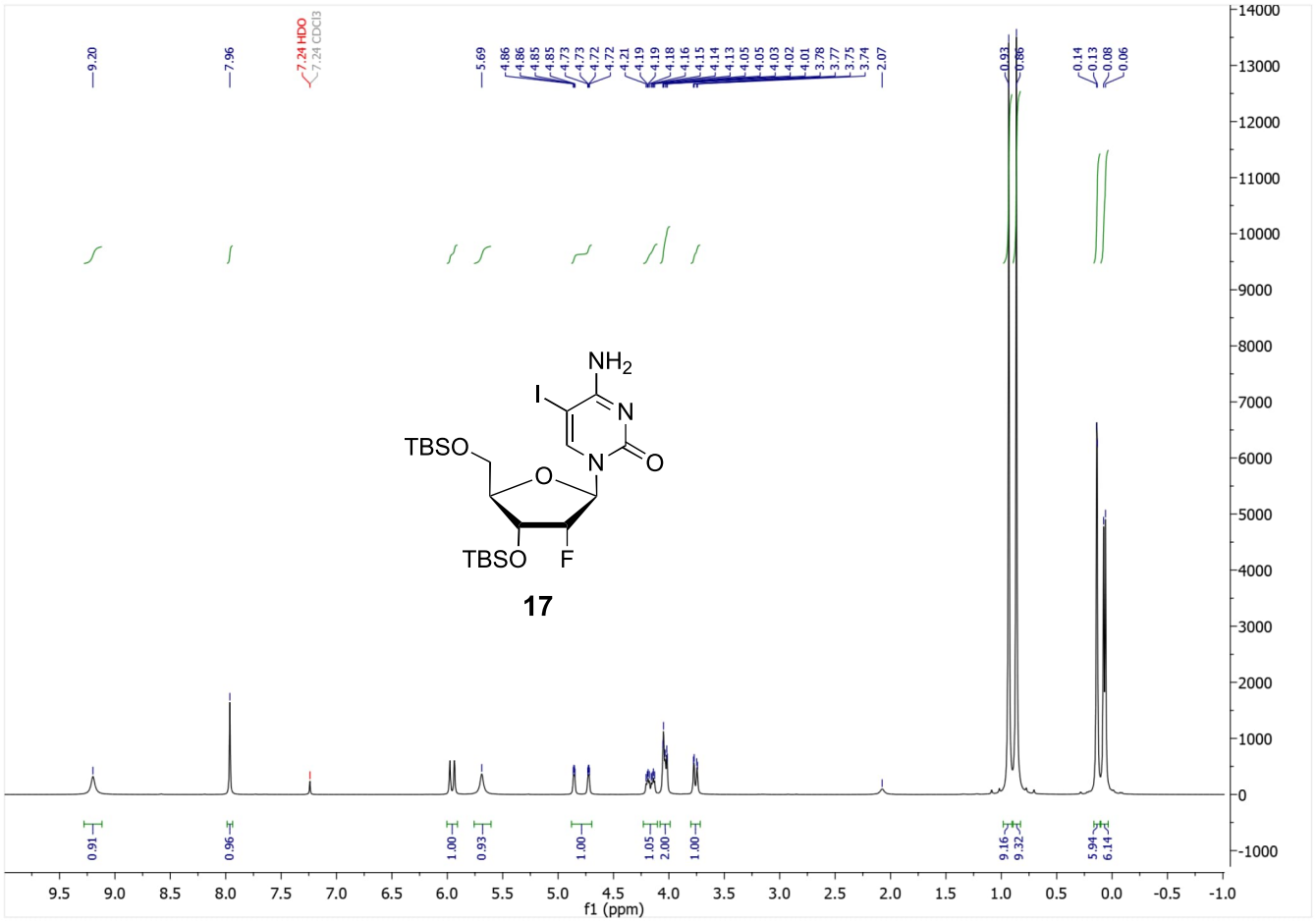


Figure SI-41: ^{1}H-NMR spectrum of compound **17**.

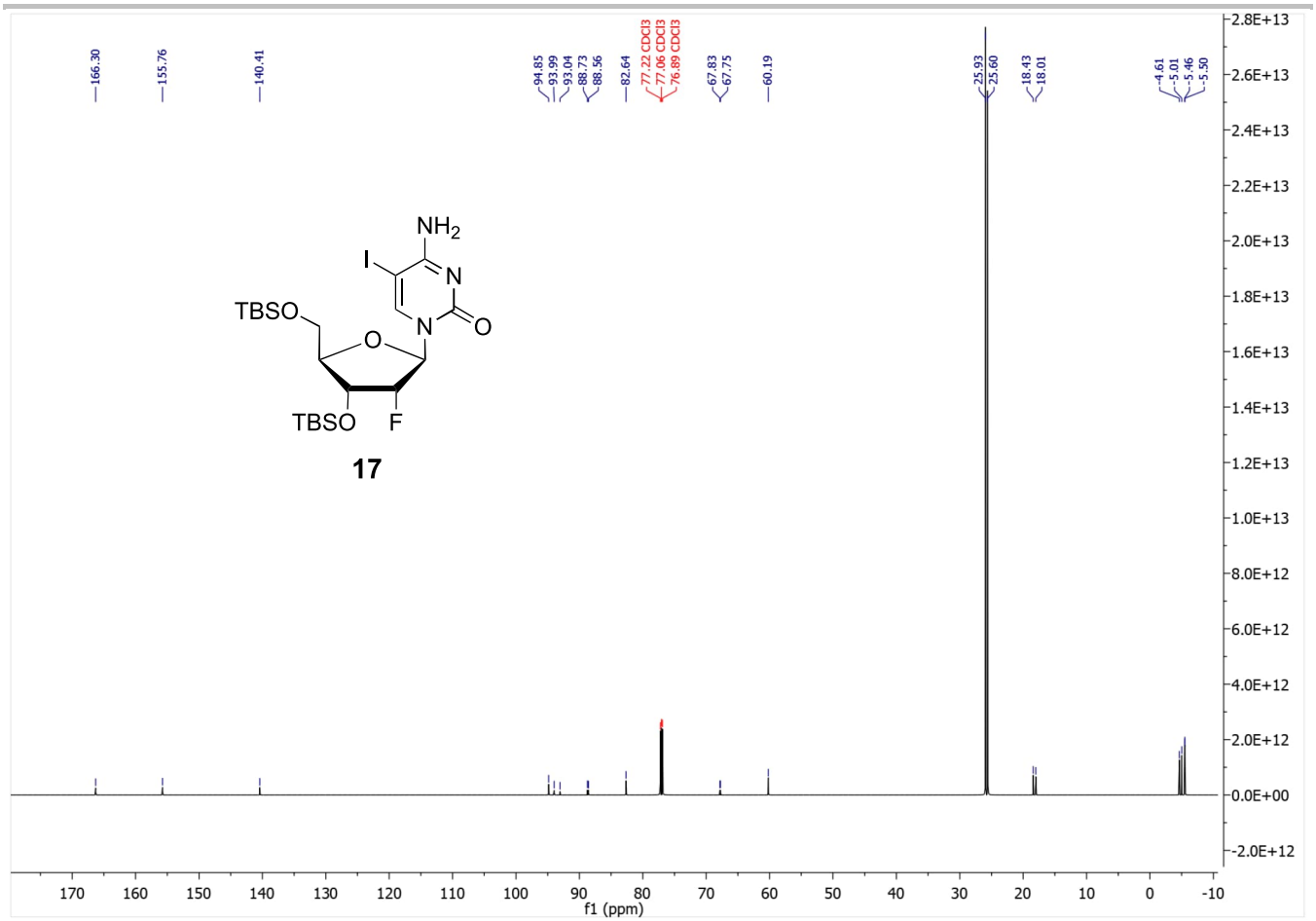


Figure SI-42: ^{13}C-NMR spectrum of compound **17**.

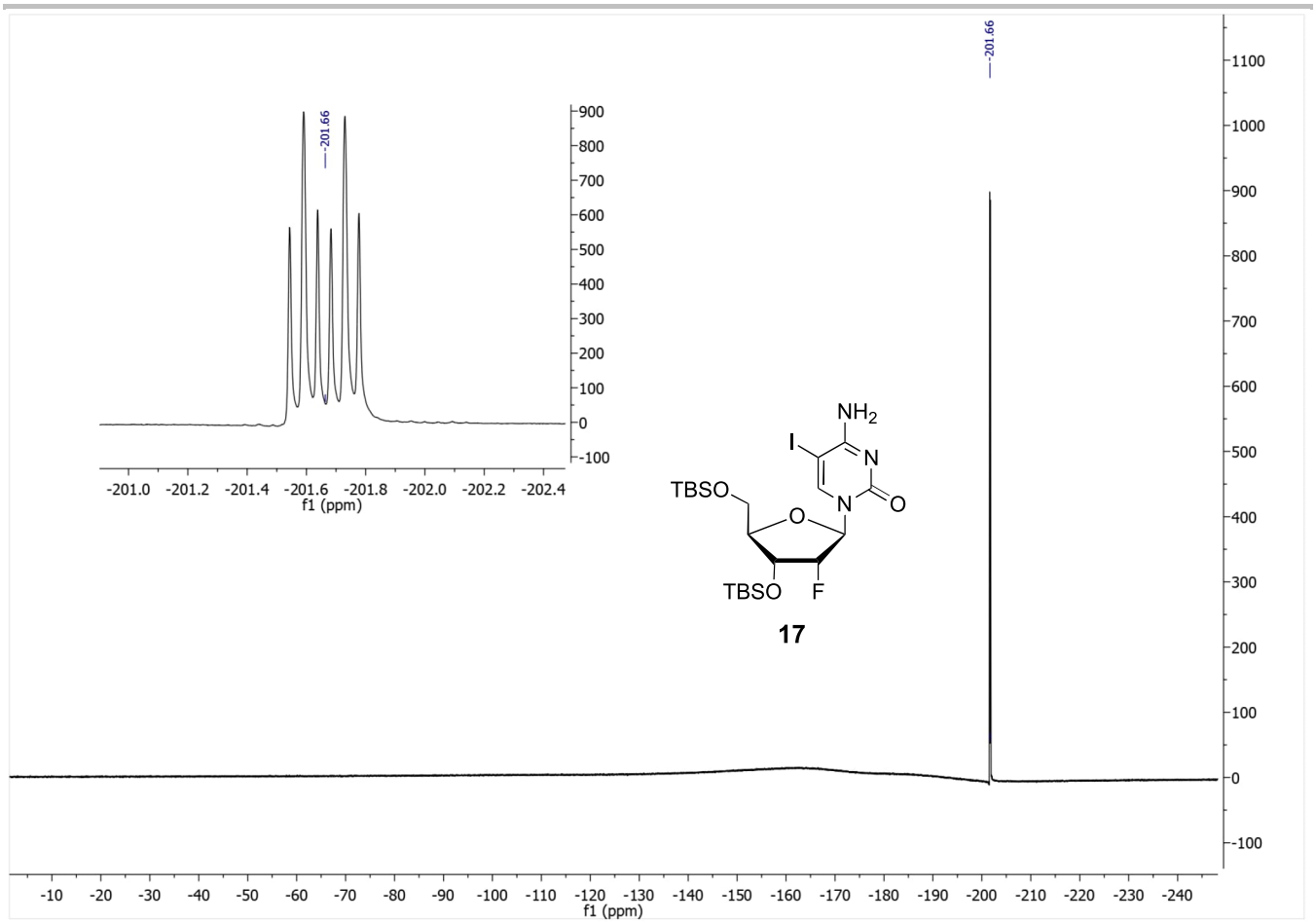


Figure SI-43: ^{19}F-NMR spectrum of compound **17**.

3'-5'-*O*-bis(*tert*-Butyldimethylsilyl)-5-methoxycarbonyl-2'-desoxy-2'-(R)-fluoro-cytidine (18)

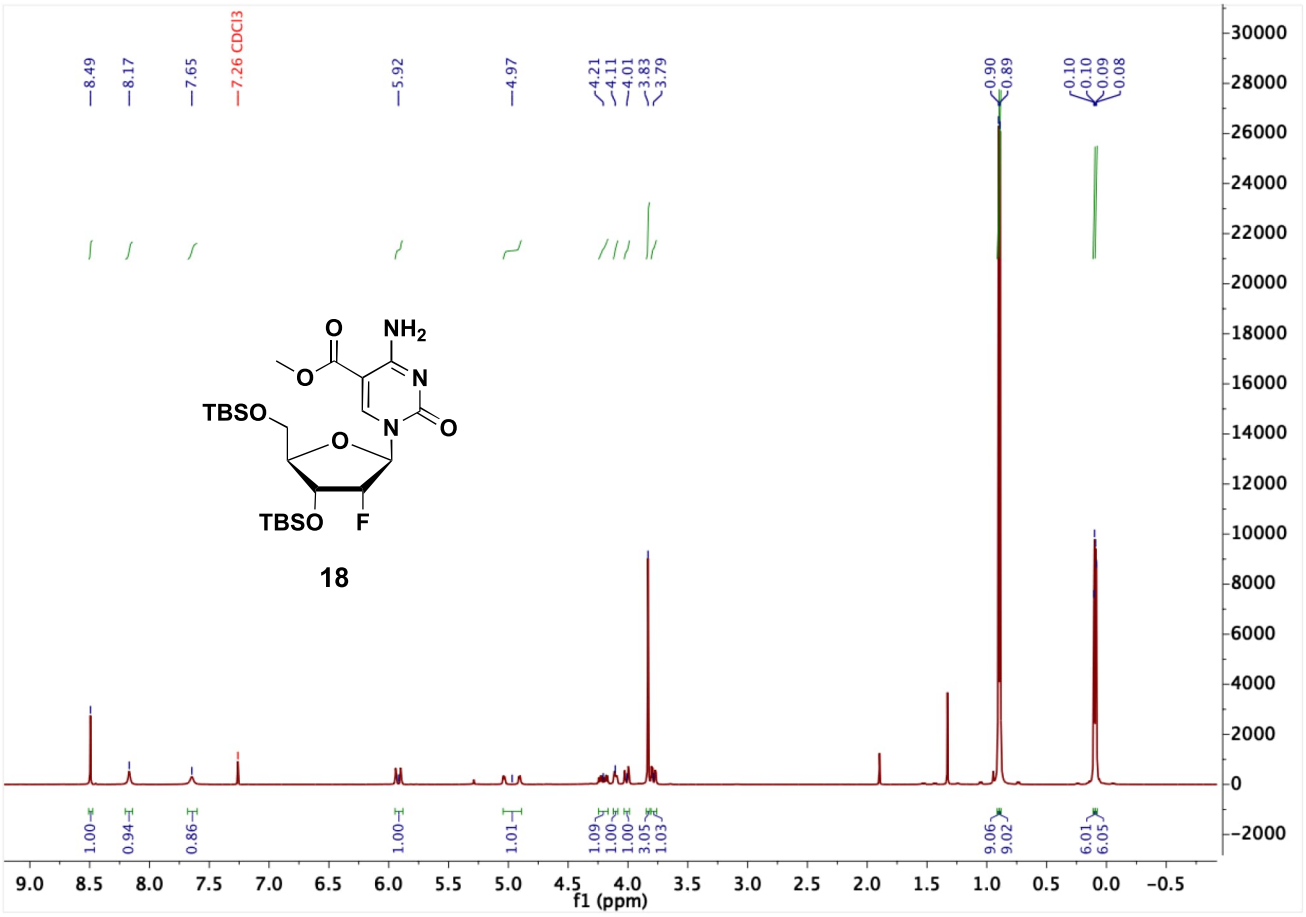


Figure SI-44: ^{1}H-NMR spectrum of compound **18**.

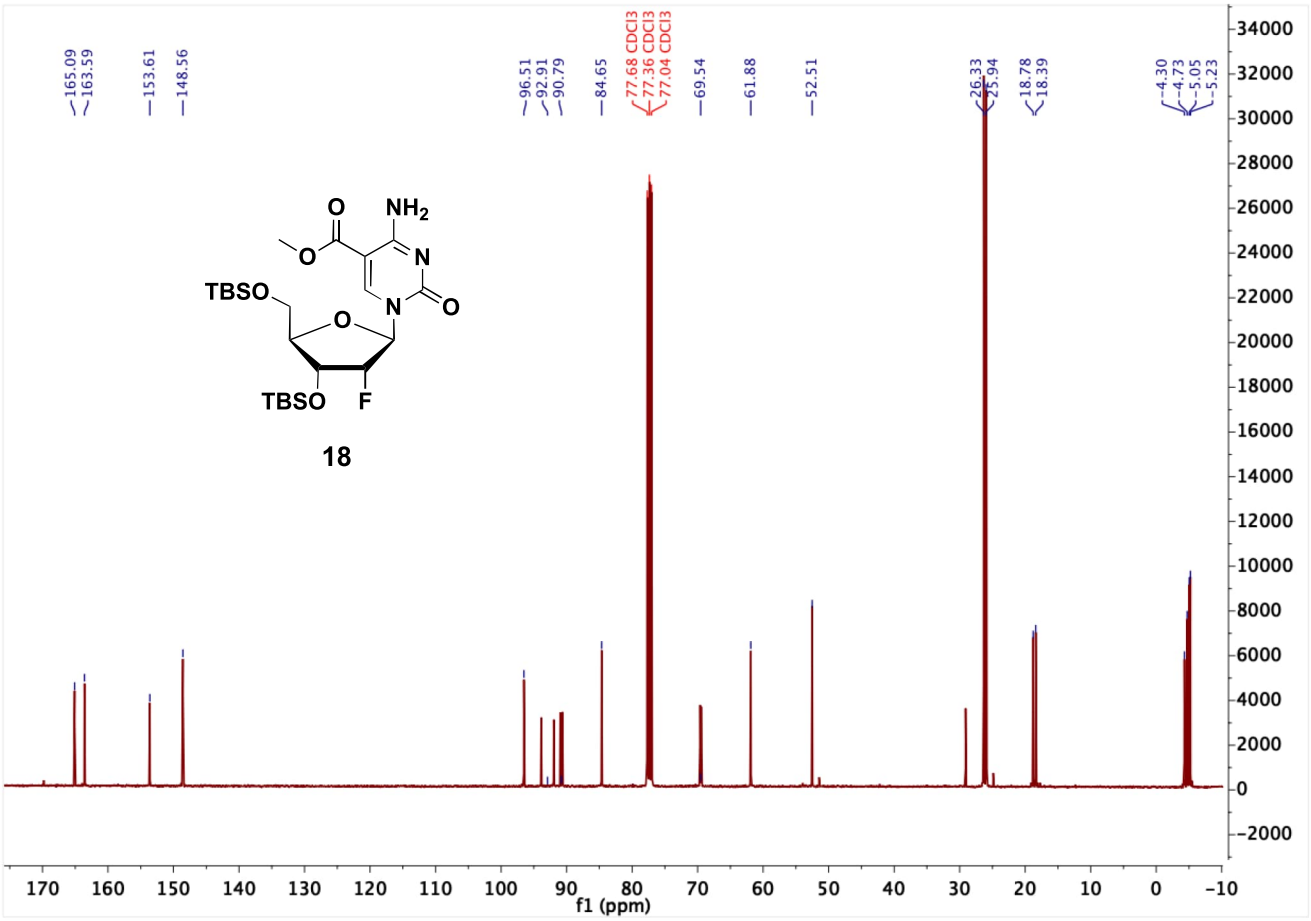


Figure SI-45: ^{13}C-NMR spectrum of compound **18**.

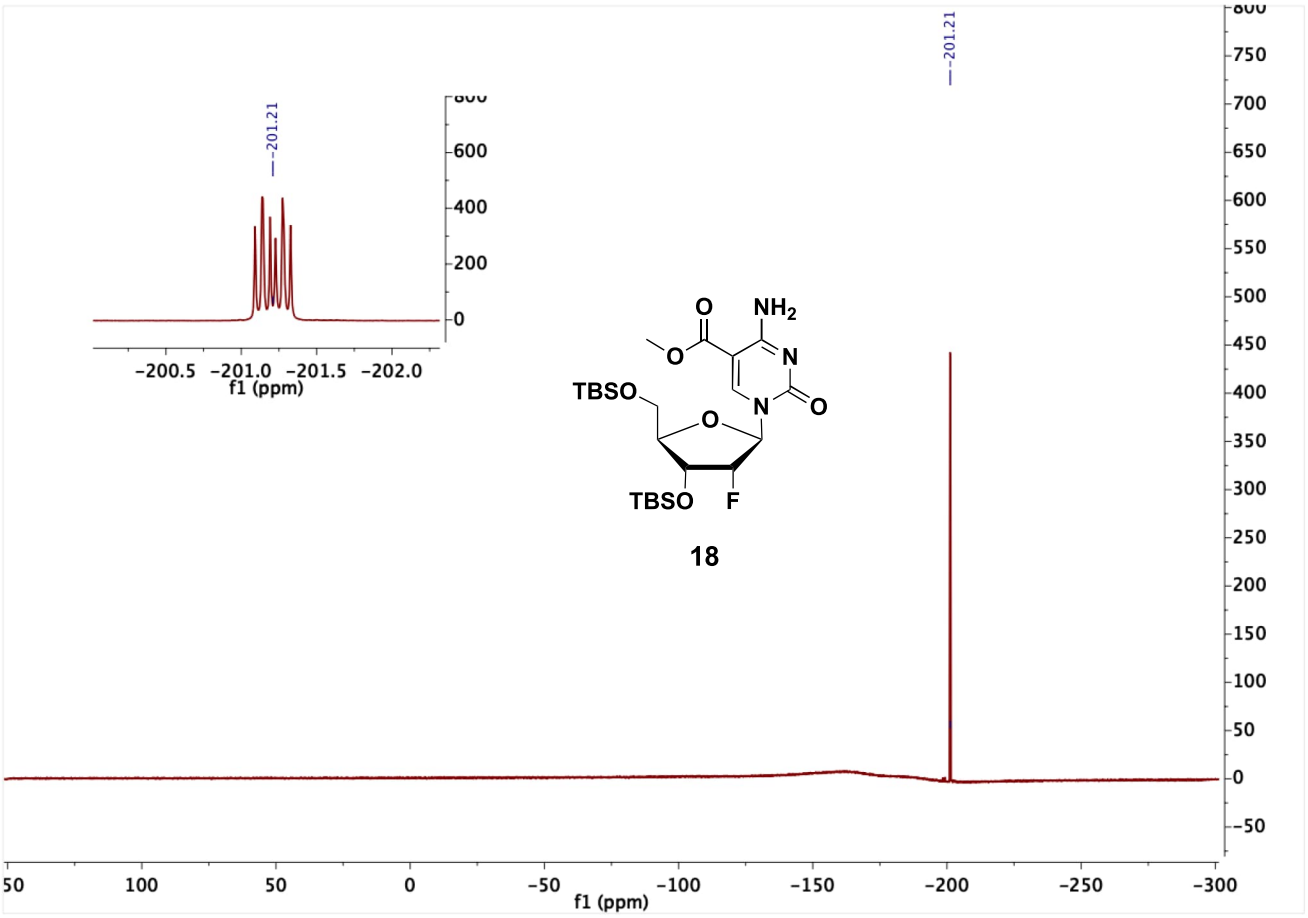


Figure SI-46: ^{19}F-NMR spectrum of compound **18**.

3'-5'-*O*-bis(*tert*-Butyldimethylsilyl)-5-methoxycarbonyl-2'-desoxy-2'-(R)-fluoro-5,6-dihydrocytidine (19)

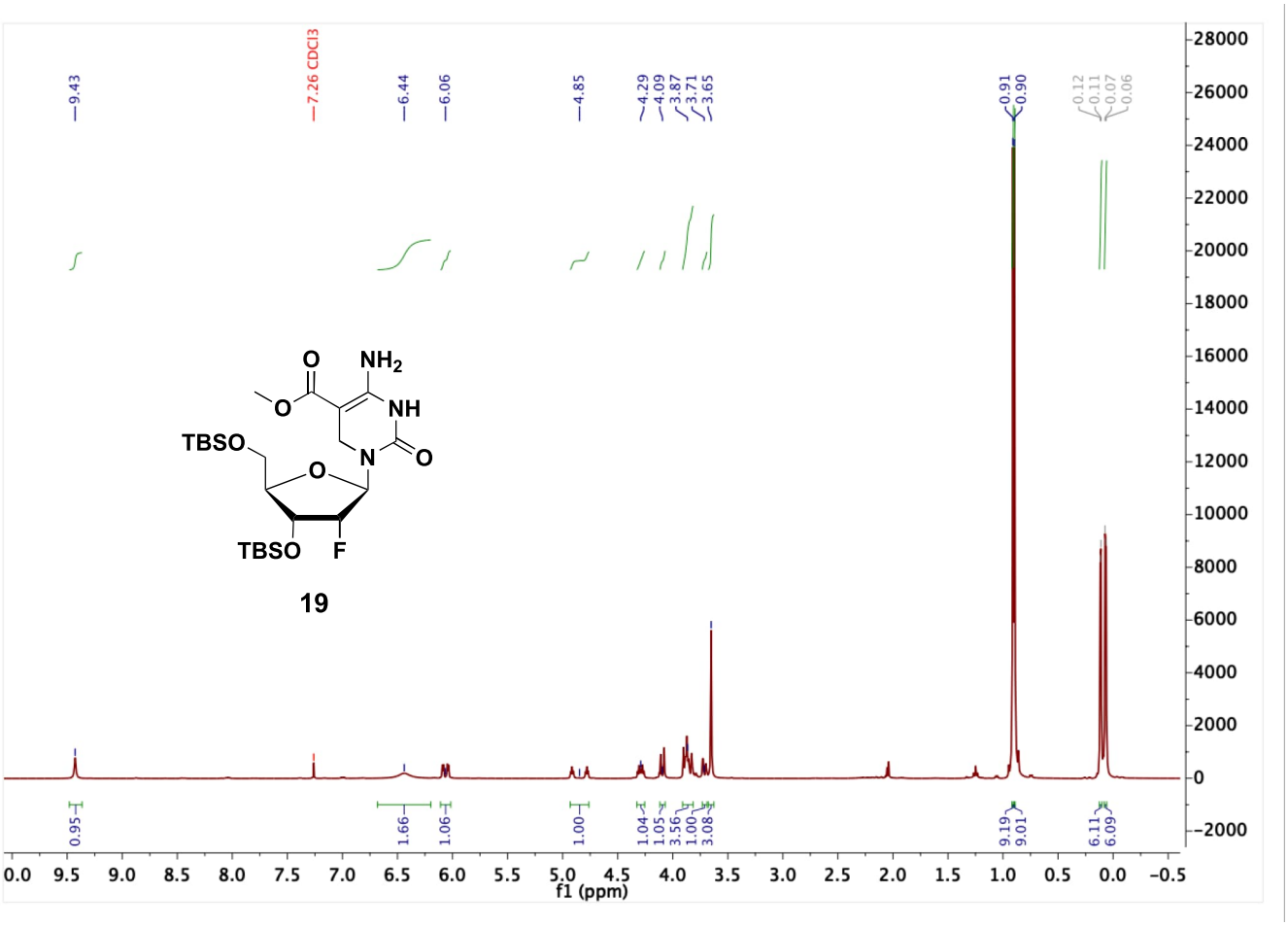


Figure SI-47: ^{1}H-NMR spectrum of compound **19**.

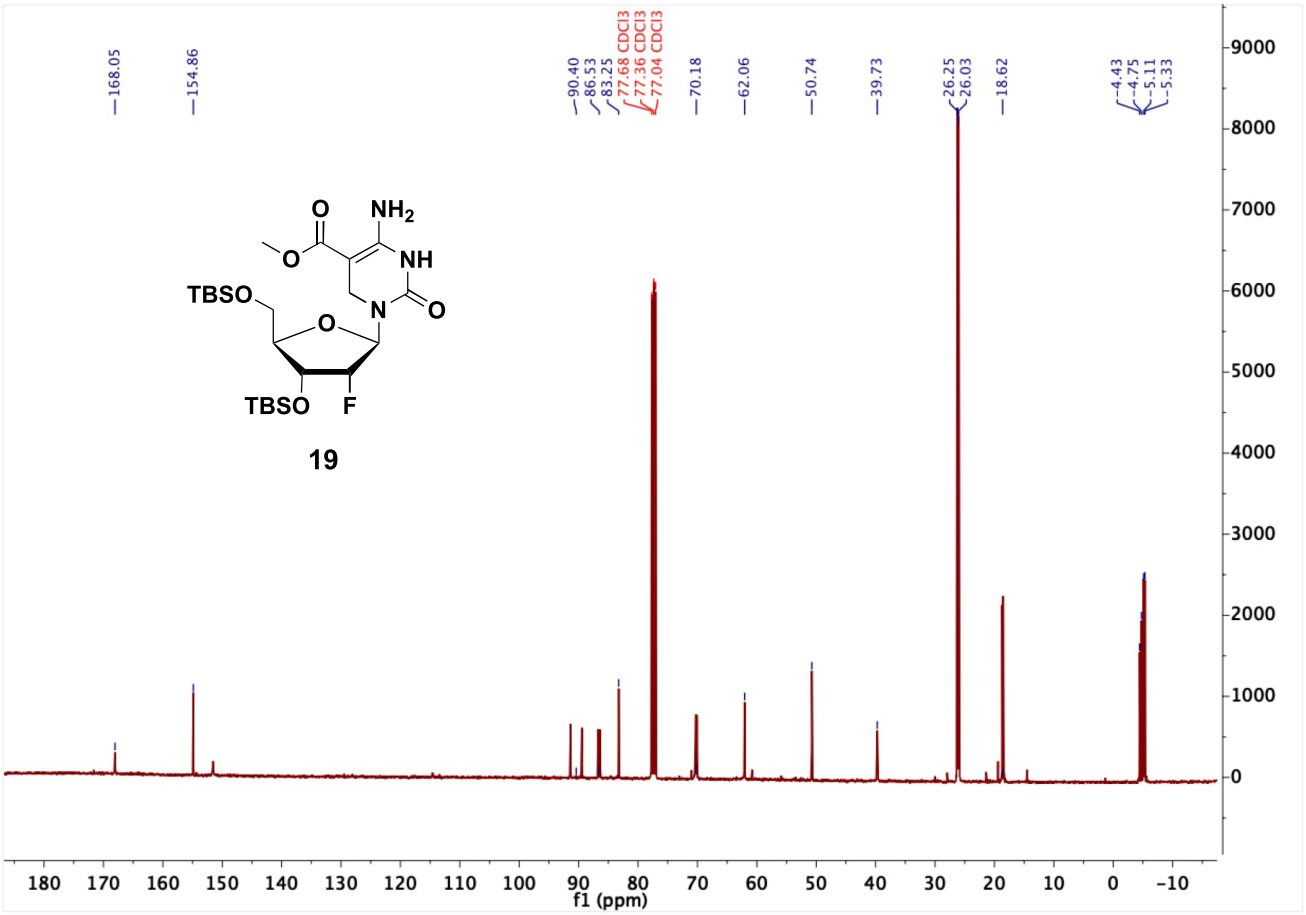


Figure SI-48: ^{13}C-NMR spectrum of compound **19**.

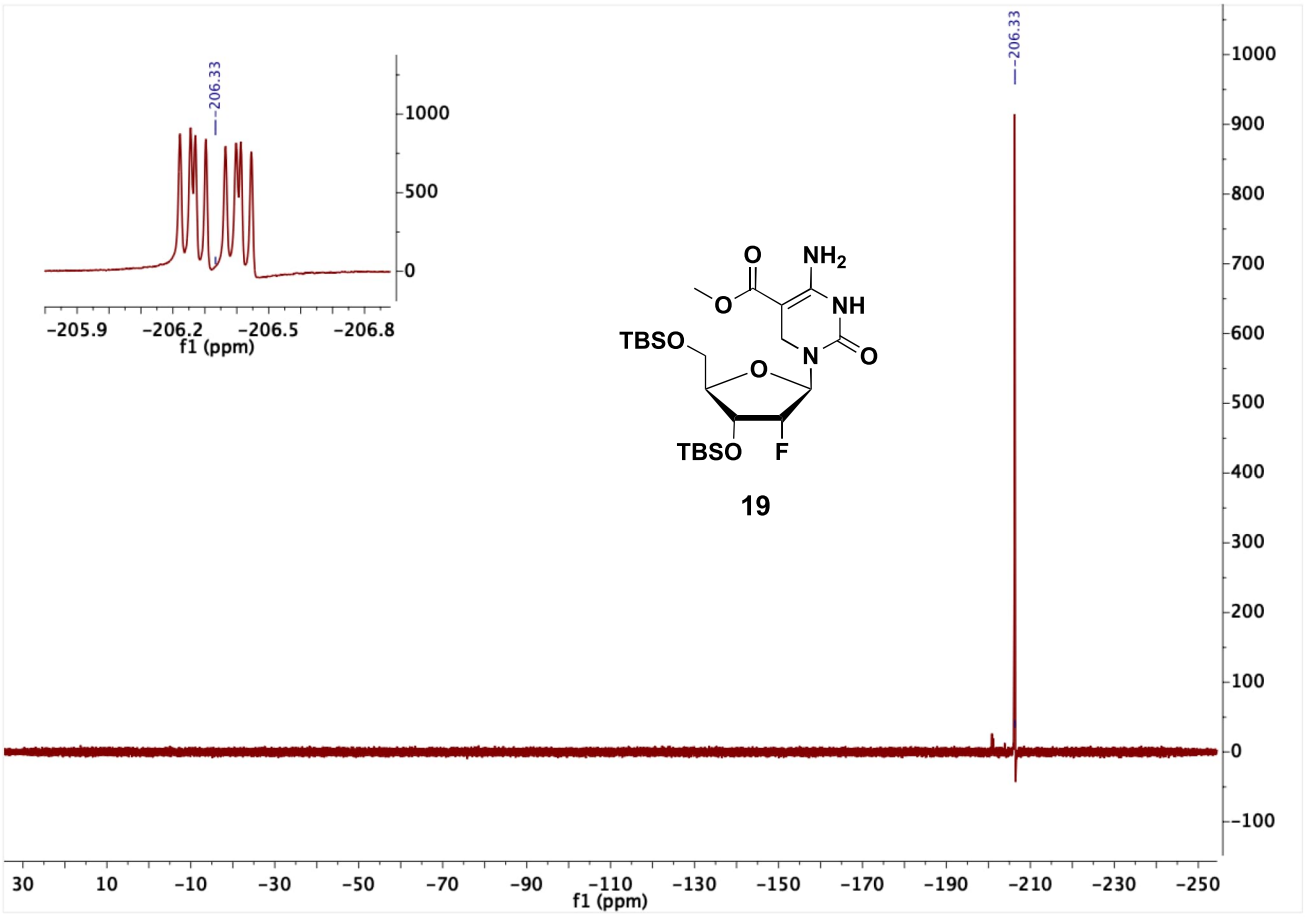


Figure SI-49: ^{19}F-NMR spectrum of compound **19**.

3'-5'-*O*-bis(*tert*-Butyldimethylsilyl)-2'-desoxy-2'-(R)-fluoro-5,6-dihydrocytidine (20)

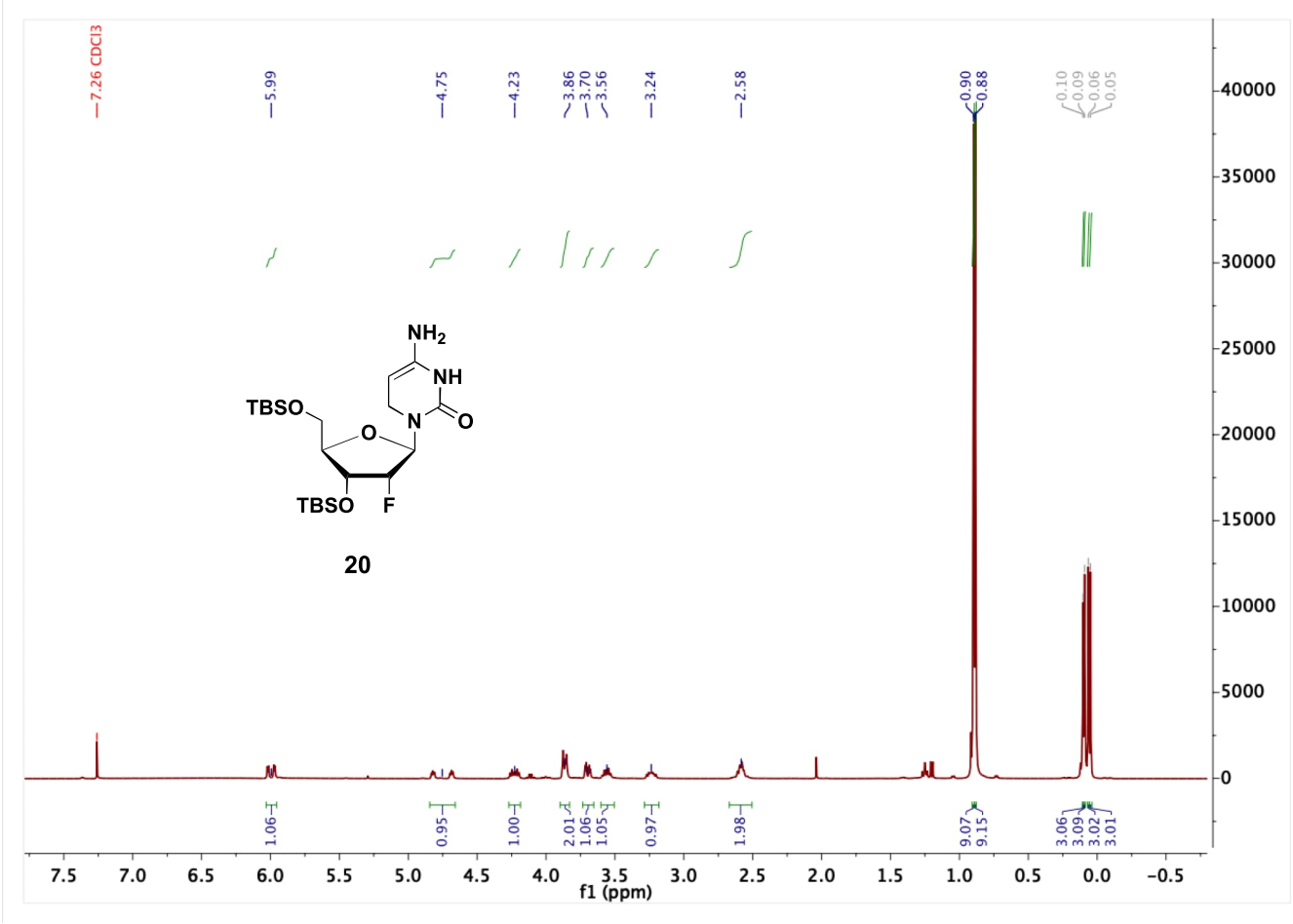


Figure SI-50: ^{1}H-NMR spectrum of compound **20**.

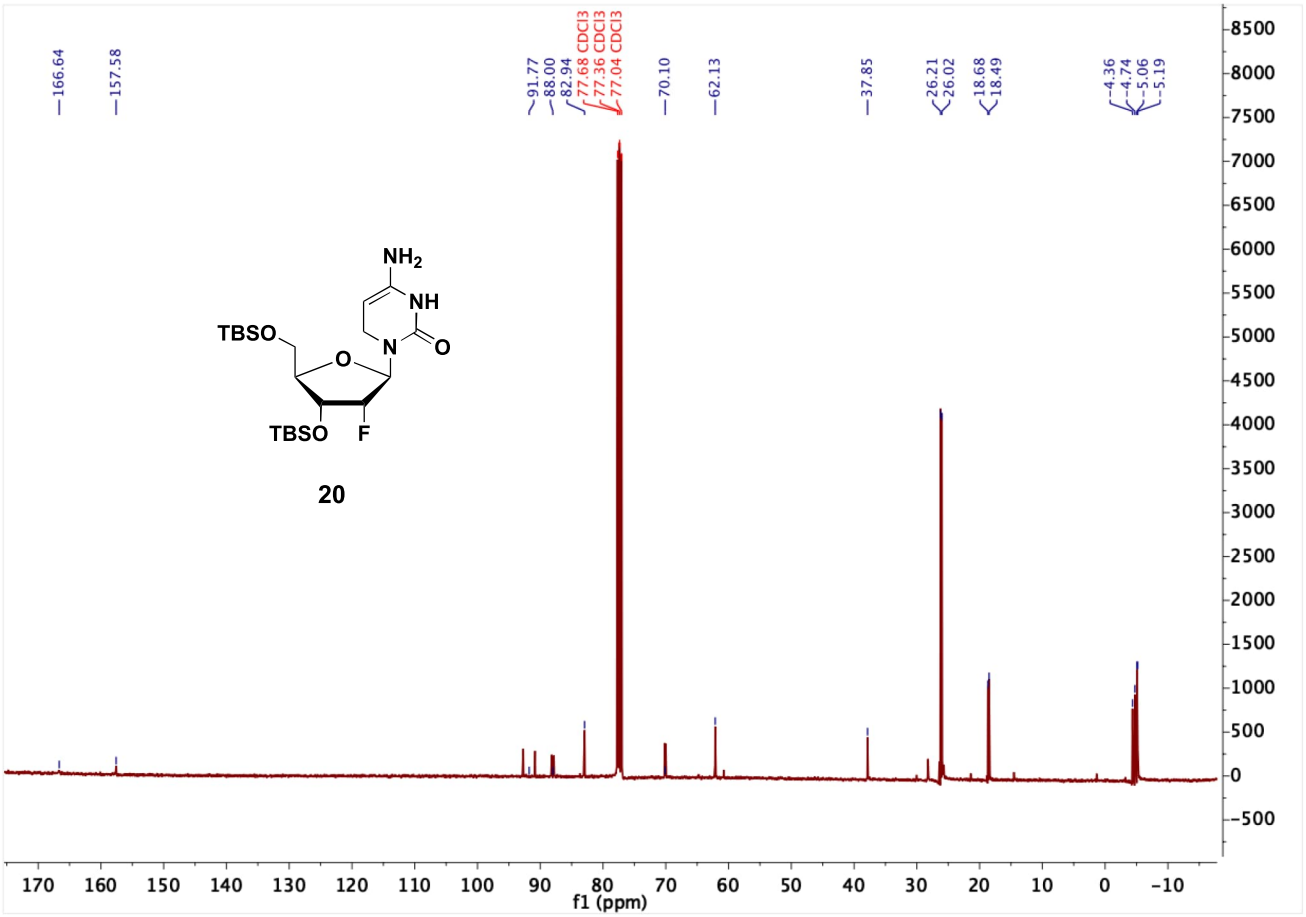


Figure SI-51: ^{13}C-NMR spectrum of compound **20**.

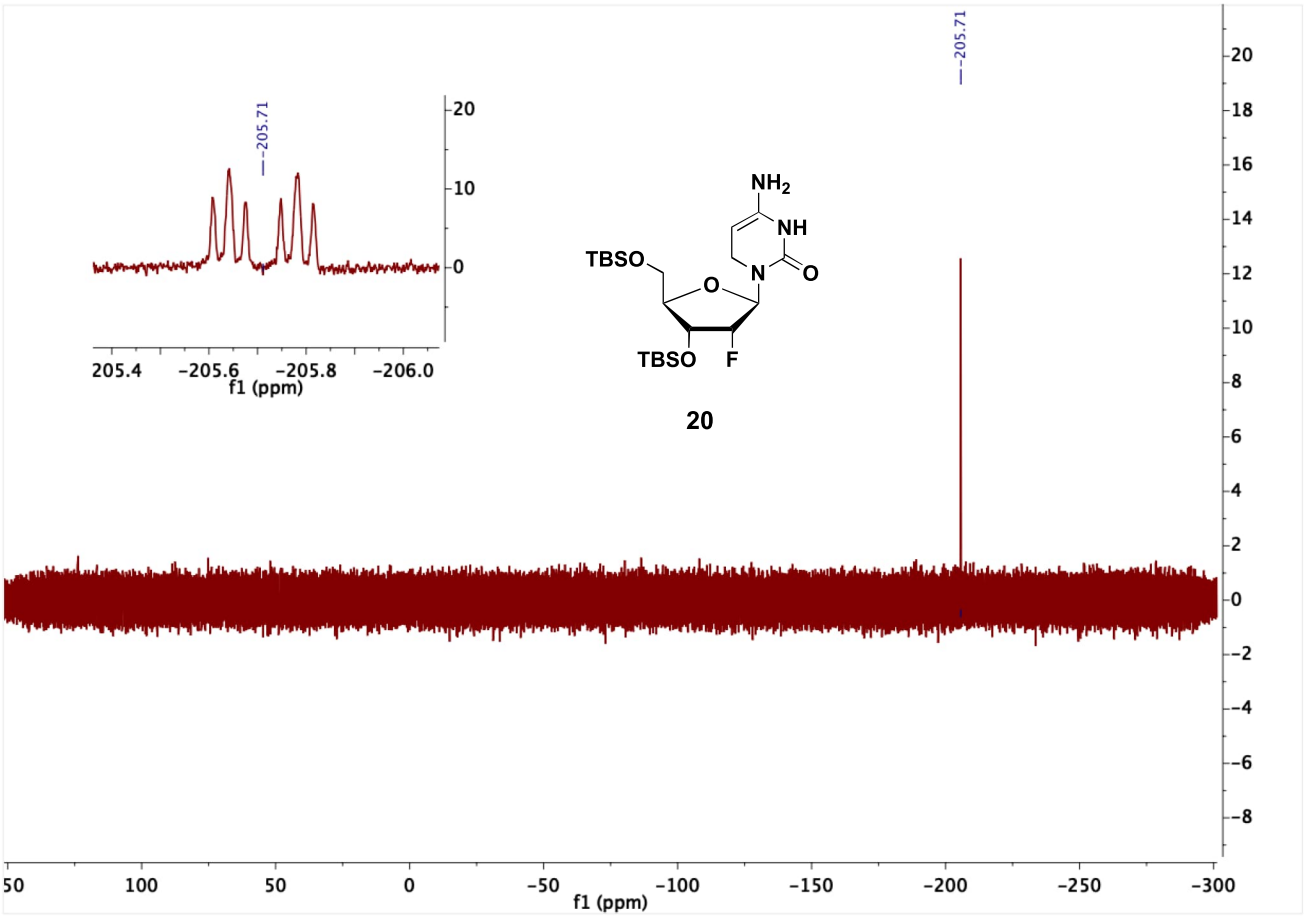


Figure SI-52: ^{19}F-NMR spectrum of compound **20**.

References

[1] F. Yuan, Y. Bi, J.-Y. Zhang, Y.-L. Zhou, X.-X. Zhang, C.-X. Song, *RSC Adv.* **2019**, *9*, 29010-29014.
[2] W. J. Lowary PT, *J Mol. Biol.* **1998**, *276*, 19-42.
[3] K. Iwan, R. Rahimoff, A. Kirchner, F. Spada, A. S. Schröder, O. Kosmatchev, S. Ferizaj, J. Steinbacher, E. Parsa, M. Müller, T. Carell, *Nat. Chem. Biol.* **2018**, *14*, 72-78.
[4] F. R. Traube, S. Schiffers, K. Iwan, S. Kellner, F. Spada, M. Müller, T. Carell, *Nat. Protoc.* **2019**, *14*, 283-312.
[5] A. S. Schröder, E. Parsa, K. Iwan, F. R. Traube, M. Wallner, S. Serdjukow, T. Carell, *Chem. Commun.* **2016**, *52*, 14361-14364.
[6] A. S. Schröder, O. Kotljarova, E. Parsa, K. Iwan, N. Raddaoui, T. Carell, *Org. Lett.* **2016**, *18*, 4368-4371.

8.3 Supporting Information: Biomimetic iron complex achieves TET enzyme reactivity

D. Schmidl*, N. Jonasson*, E. Korytiaková*, T. Carell, L. Daumann *Angew. Chem. Int. Ed.* **2021**,*60*,21457–21463.

*: Diese Autoren haben zu gleichen Teilen beigetragen

Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

Biomimetic Iron Complex Achieves TET Enzyme Reactivity**

*David Schmidl[+], Niko S. W. Jonasson[+], Eva Korytiaková[+], Thomas Carell, and Lena J. Daumann**

anie_202107277_sm_miscellaneous_information.pdf

Supporting Information
©Wiley-VCH 2019
69451 Weinheim, Germany

Table of Contents

1. Materials and Methods
1.1. Solvents and chemicals
1.2. Methods and manipulations
1.3. Gas chromatography-mass spectroscopy (GC-MS)
1.4. NMR spectroscopy
1.5. ESI and EI Mass spectrometry
1.6. Nucleoside Quantification in Oligonucleotide Experiments
1.7. High performance liquid chromatography (HPLC) in Nucleoside Experiments
1.8. UV/Vis Oligonucleotide Quantification
1.9. Matrix assisted laser desorption ionization mass spectrometry (MALDI MS)
1.10. Bond Dissociation Energies and Measured Kinetic Constants
Experimental Section
2. Nucleoside Experiments
2.1. General Procedure for HPLC Experiments
2.2. HPLC Traces
2.2.1. Overview
2.2.2. HPLC trace of 5mdC, 5hmdC, 5fdC, and 5cadC
2.2.3. HPLC trace of 5mC, 5hmC, 5fC, and 5caC
2.2.4. HPLC trace of 1 after filtration
2.2.5. Spike Experiments – No Additions
2.2.6. Spike Experiments – 5mdC or 5mC
2.2.7. Spike Experiments – 5hmdC or 5hmC
2.2.8. Spike Experiments – 5fdC or 5fC
2.2.9. Spike Experiments – 5cadC or 5caC
2.3. GC-MS Reference Samples
2.3.1. 2'-Deoxyribono-γ-lactone
2.3.2. 5-methyl-2'deoxyribo-cytidine
2.3.3. Overlay of References
2.4. Reactions of 1 with substrates
2.4.1. Reaction of 1 with 5mdC
2.4.2. Reaction of 1 with 5hmdC
2.4.3. Reaction of 1 with d_3-5mdC
2.4.4. Filtration and control experiments
3. Oligonucleotide Experiments
3.1. General Procedure for Oligonucleotide Experiments
3.2. Oligonucleotide Concentration

3.3. Digestion and Quantification Procedure
3.4. Quantification Results
3.5. Product Ratios
3.6. Side-Product Analysis
4. Synthetic Procedures
4.1. Synthesis of 1
4.2. Synthesis of 5-formyl-2'deoxy-cytidine (5fdC)
4.2.1. Synthesis of silyl protected 2'-deoxy-cytidine (2)
4.2.2. Synthesis of silyl protected 5-iodo-2'-deoxy-cytidine (3)
4.2.3. Synthesis of silyl protected 5-formyl-2'-deoxy-cytidine (4)
4.2.4. Synthesis of 5-formyl-2'deoxyx-cytidine (5fdC)
4.3. Synthesis of 5cadC
4.3.1. Synthesis of silyl protected 5'-carboxyethyl-2'-deoxy-cytidine (5)
4.3.2. Synthesis of 5-carboxy-2'-deoxycytidine (5cadC)
4.4. Synthesis of nucleobases 5-formylcytosine (5fC) and 5-carboxycytosine (5caC)
4.4.1. Synthesis of 5-formylcytosine (5fC)
4.4.2. Synthesis of 5caC
4.5. Synthesis of D_3-5mdC
5. Appendix
5.1. NMR spectra
5.1.1. TBS-dC 2
5.1.2. 5-Iodo-TBS-dC 3
5.1.3. 5-Formyl-TBS-dC 4
5.1.4. 5-Formyl-2'-deoxycytidine (5fdC)
5.1.5. 5-(Methoxycarbonyl)-TBS-dC 5
5.1.6. 5-carboxyl-2'-deoxycytidine (5cadC)
5.1.7. 5-Formylcytosine (5fC)
5.1.8. 5-carboxylcytosine (5caC)
5.1.9. D_3-5-methyl-2'-deoxycytidine (D_3-5mdC)
References

1. Materials and Methods

1.1. Solvents and chemicals

Chemicals were purchased from commercial sources (Sigma Aldrich, ABCR, Acros Organics, Alfa Aesar, TCI Chemicals, Oakwood Chemic used without further purification except for dichloromethane, diethyl ether and hexanes. The solvents were purchased from the LMU Mun reduced pressure prior to use.

1.2. Methods and manipulations

All manipulations were carried out in flame-dried glassware under a positive pressure of nitrogen gas and magnetically stirred if not stated oth and absolute solvents were transferred *via* stainless-steel cannula or syringe. Organic solutions were concentrated by rotary evaporation at (TLC) was performed on pre-coated (silica gel, 0.25 mm, 60 Å pore-size, 230–400 mesh, Merck KGA) aluminum plates or which were impre TLC plates were visualized by exposure to ultraviolet light.

1.3. Gas chromatography-mass spectroscopy (GC-MS)

Reference samples were used as received, reaction samples were dried under high vacuum (<1*10^{-2} mbar) prior to derivatization. 0.5 m acetonitrile, 100 µl BSTFA were added and the mixture heated to 70 °C for 30 min. The samples were filtered and injected (1 µl, split or splitless onto an Agilent® 7920 GC equipped with a 30 m HP5-MS column (Agilent® 19091S-433UI) coupled to an Agilent® 5970 EI mass spectr 280 °C and the temperature of the ion source 230 °C. The initial oven temperature was 80 °C, held there for 2 min, ramped to 240 °C at 5 spectra were recorded in scan mode between 70-400 m/z.

1.4. NMR spectroscopy

^{1}H NMR, ^{2}H NMR and ^{13}C NMR spectra were recorded at room temperature on Bruker Avance III (400MHz) operating at 400MHz for prot 100MHz for carbon nuclei. ^{1}H-chemical shifts are reported in ppm units relative to $CDCl_3$ (δ_H = 7.26), CD_3CN (δ_H = 1.94), or D_2O (δ_H = 4.79 relative to $CDCl_3$ (δ_C = 77.16), CD_3CN (δ_C = 1.32) or DMSO-d_6 (δ_C = 39.52).[1] The following abbreviations were used: s = singlet, d = dt = doublet of triplets, m = multiplet, br = broad. Coupling constants (*J*) are given in Hertz. The software used for data processing was MNo

1.5. ESI and EI Mass spectrometry

ESI spectra were recorded with a Thermo Finnigan LTQ FT Ultra Fourier Transform Ion Cyclotron Resonance mass spectrometer with aceto Quantitative UHPLC-MS/MS analysis of digested DNA samples was performed using an Agilent 1290 UHPLC system equipped with a UV d mass spectrometer. The source-dependent parameters were as follows: gas temperature 80 °C, gas flow 15 L/min (N_2), nebulizer 30 psi, s L/min (N2), capillary voltage 2.500 V in the positive ion mode, capillary voltage −2.250 V in the negative ion mode and nozzle voltage 500 V Delta EMV was set to 500 (positive mode) and 800 (negative mode). Compound-dependent parameters are summarized in Supplementary T a Poroshell 120 SB-C8 column (Agilent, 2.7 µm, 2.1 mm × 150 mm) at 35 °C using a gradient of water and MeCN, each containing 0.0 mL/min: 0 →4 min; 0 →3.5% (v/v) MeCN; 4 →7.9 min; 3.5 →5% MeCN; 7.9 →8.2 min; 5 →80% MeCN; 8.2 →11.5 min; 80% MeCN; 11.5 MeCN. The autosampler was cooled to 4 °C. The injection volume was amounted to 39 µL.

1.6. Nucleoside Quantification in Oligonucleotide Experiments

In the case of the oligonucleotide experiments, nucleosides of interest were quantified using the stable isotope dilution technique.[2] Internal sta 5fdC*, 5cadC*, and 8-oxo-dG*) were added to a sample and the samples were then digested using Degradase Plus from Zymo Researc UHPLC-MS/MS system (UHPLC system: Agilent Technologies, model no. 1290 Infinity II LC; MS system: Triple quadrupole LC/MS system w model no. 6490)). The exact method was described here.[2] The sample data were analyzed by Quantitative MassHunter Software from Agi using the signal of the UV-trace with known calibration curves.[2] Used internal calibration curves, as well as UV-calibration curves are shown

Table S1: Calibration curves for UHPLC-MS/MS quantification of nucleosides after enzymatic digestion of the oligonucleotide product mixture.

Nucleoside	Linear Regression
UV-dC	y = 0.28558x – 0.02812
UV-dG	y = 0.54214x + 0.1554
mdC	y = 0.75157x – 3.70649*10^{-4}
hmdC	y = 0.89129 – 3.96*10^{-3}
fdC	y = 1.18055x – 0.00206
cadC	y = 0.94585x – 0.01218
8-oxo-dG	y = 0.9355 – 0.00138

1.7. High performance liquid chromatography (HPLC) in Nucleoside Experiments

HPLC traces were recorded on two Agilent 1260 Infinity II HPLC systems, each equipped with a flexible quaternary pump (G7104C), a via (G7129C) and a multi-wavelength detector (G7165A). An ACE C-18/PFP column (150×4.6mm; 5 µm; 100 Å) was used with a binary solver A/1% B over 18 min (A: 0.1% TFA in H_2O; B: 0.1% TFA in MeCN). Quantifications were performed on both instruments, the respective calculations are given below **(Table S2, Figure S1).**

Table S2: Calibration curves for UV absorption quantification in HPLC traces. Two different instruments were used for the experiments with **1** and 5mdC (HPLC1) and for

Substrate	**Linear Regression**
HPLC 1	
5mdC	y = 445.99x – 15.244
5hmdC	y = 451.86x – 3.3986
5fdC	y = 550.06x – 14.941
5cadC	y = 289.74x – 5.9377
HPLC 2	
5mdC	y = 860.96x + 78.134
5hmdC	y = 905.75x + 85.083
5fdC	y = 1016.5x + 89.150
5cadC	y = 611.05x + 55.824
5mC	y = 609.33x + 41.662
5hmC	y = 601.37x + 41.525

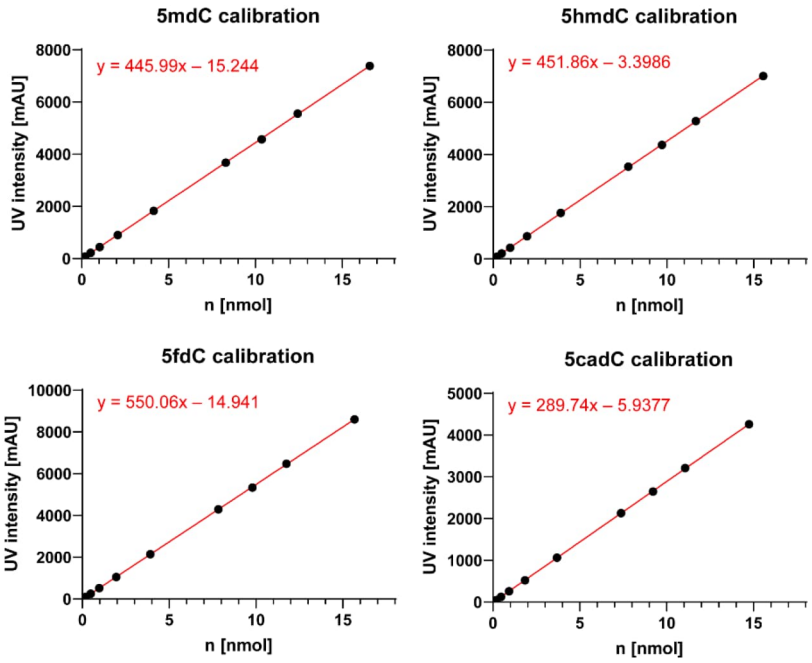


Figure S1: HPLC (UV, λ = 280 nm) Quantification curves for the nucleosides 5mdC, 5hmdC, 5fdC and 5cadC in the nucleoside experiments.

Table S3: Proportionality factors for the quantification of nucleobases and D_3-5mdC in the nucleobase experiments based on the nucleoside standard calibrations given in

Nucleobase/Nucleoside	Proportionality Factor
5mC/5mdC	0.692915425
5hmC/5hmdC	0.652580968
5fC/5fdC	0.606517414
5caC/5cadC	0.409907652
D_3-5mdC/5mdC	0.755318881

1.8. UV/Vis Oligonucleotide Quantification

The concentration of oligonucleotide reaction samples were measured after filtration, lyophilisation and re-suspension. The absorption at (NanoDrop Technologies, cat. no. ND-1000) and the concentration determined by the following equation (***Equation 1***).

$$c = \frac{A_{260}}{\varepsilon_{260}} \qquad \textit{Equation 1}$$

$c\ [mM]$ is the concentration, $A_{260}\ [cm]$ is the absorption measured at λ = 260 nm corrected for d = 1 cm path length, $\varepsilon_{260}\ [mM^{-1}cm^{-1}]$ oligonucleotide (as supplied by the manufacturer Sigma Aldrich).

Table S4: Abbreviations for the used oligonucleotide strands and corresponding absorption coefficients at λ = 260 nm.

Strand sequence (5'->3')	Abbreviation	ε_{260} [mM^{-1} cm^{-1}
HO-CCUUAACC5mdCG-OH	**O**	**90**
HO-CCUUAACCCG-OH	**No5mC**	**87**
HO_3PO-CCUUAACCCG-OH	**No5mC-P**	**87**

1.9. Matrix assisted laser desorption ionization mass spectrometry (MALDI MS)

MALDI mass spectra were recorded on a Bruker Daltonics Autoflex II time-of-flight mass spectrometer equipped with a nitrogen laser (λ = 3 matrix substance. The acceptable deviation of the instrument is $\pm$3 m/z and peaks are observed with peak widths of 5-15 m/z. Therefore, was deemed to be 10-20 m/z. It should be noted that this means that cytosine/cytidine cannot be differentiated from uracil/uridine. Diff adenine/adenosine should be possible.

1.10. Bond Dissociation Energies and Measured Kinetic Constants

Table S5: Calculated bond dissociation energies of various relevant carbon atoms in nucleoside substrates and rate constants previously measured for the reaction of 5m or in DNA context (using TET2).

Hydrogen Atom[a]	BDE [kcal mol^{-1}][b]	k_S (**1**) [L mol^{-1} s^{-1}](c)	k_{cat} (TET2) [10^{-3} s^{-1}][d]
5mC (C**H**$_3$)	90.4	7.37	2.12
5hmC (C**H**$_2$OH)	86.2	63.56	0.63
5fC (C**H**O)	92.9	12.81	0.46
Deoxyribose 1'	87.7	-	-
Deoxyribose 5'	92.4	-	-

[a] The indicated hydrogen atom refers to the free nucleobase in the case of 5mC, 5hmC, and 5fC. For deoxyribose the relevant hydrogens atoms were calculated for dec *et al.*[3]; deoxyribose BDEs calculated by Steenken *et al.*[4] [c] Measured by Jonasson and Daumann [d] Measured by Hu *et al.*[3]

Experimental Section

2. Nucleoside Experiments

2.1. General Procedure for HPLC Experiments

Reactions of the respective nucleoside substrate with **1** were conducted in water at room temperature. The iron(IV)-oxo complex [Fe^{IV}(O)(Py_5Me_2H)][C and Daumann to remove the cerium counter ion to obtain **1** and then diluted to reach a final concentration c = 10 mM.[5] This solution was added to a reflecting the envisioned stoichiometric ratio (final concentration of **1**: 5 mM), whereupon the time count was started. After appropriate time points filtered through silica (2.0–2.5 mL) to stop the reaction, and the nucleoside and nucleobase products were eluted with 7.0 mL water. The eluted s resuspended in water. The suspension was then centrifuged (4000 rpm, 2 min) and the supernatant subjected to HPLC analysis.

2.2. HPLC Traces

2.2.1. Overview

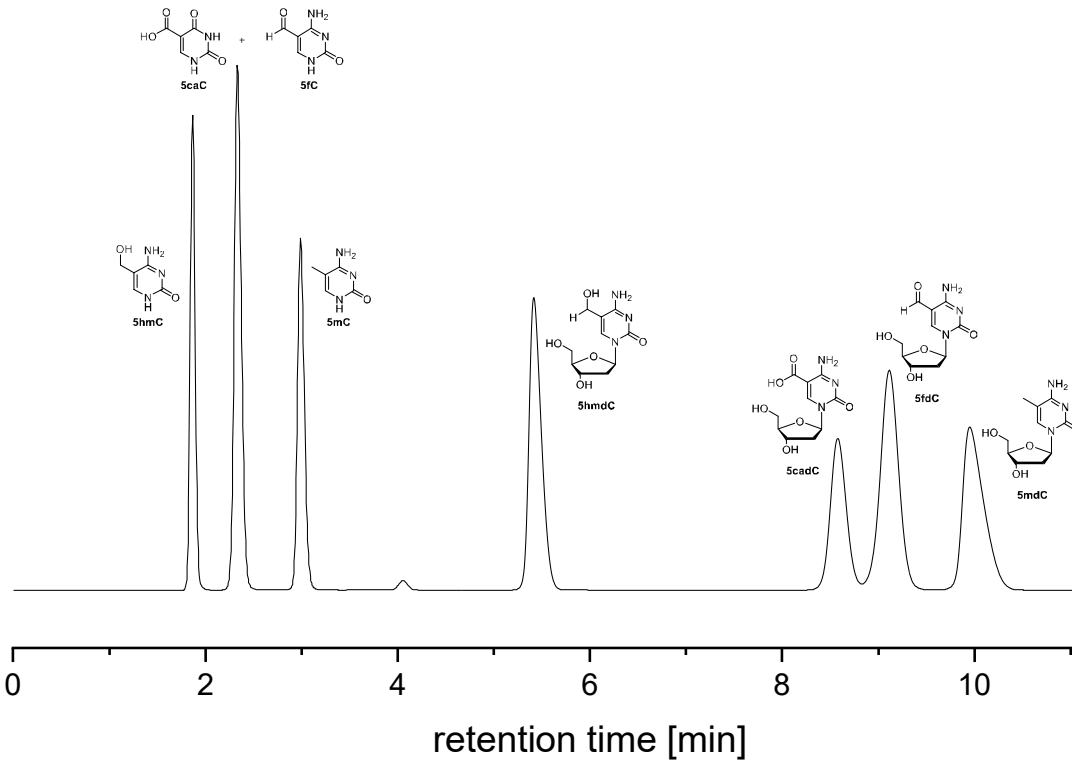


Figure S2. HPLC references traces of a mixture of 5mC, 5hmC, 5fC, 5cC, 5mdC, 5hmdC, 5fdC, and 5cadC, used for calibration purposes and shown here as refer found in 5hmdC.

2.2.2. HPLC trace of 5mdC, 5hmdC, 5fdC, and 5cadC

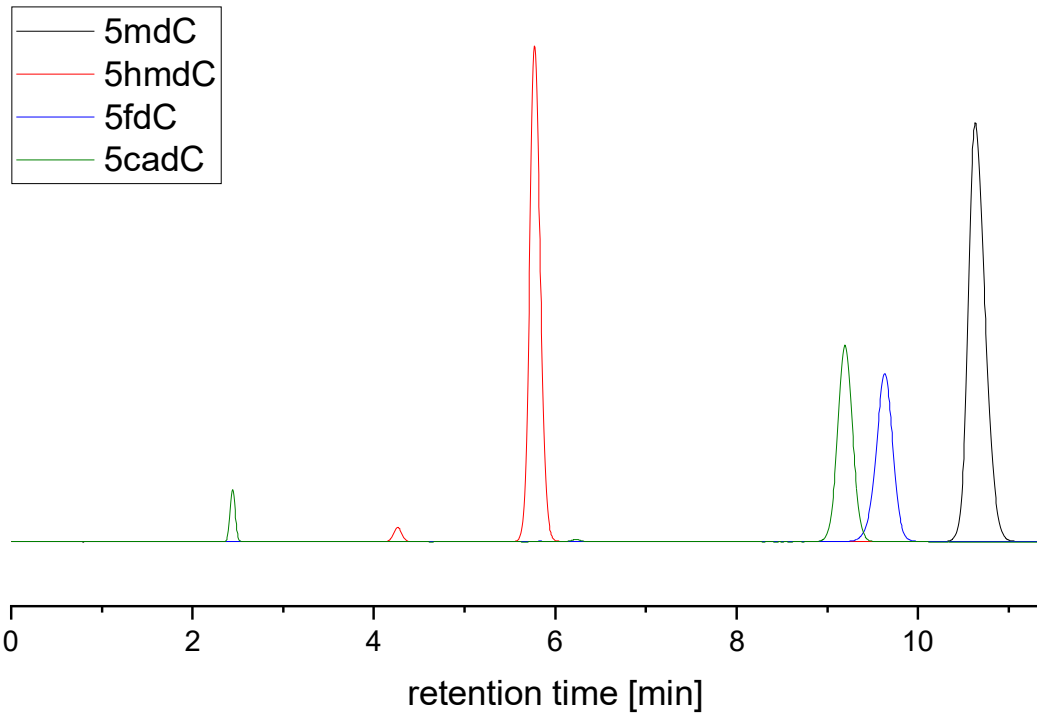


Figure S3. Superimposed HPLC references traces of 5mdC, 5hmdC, 5fdC, and 5cadC. The small signal at 4.1 min is an impurity found in 5hmdC, the small signal 5cadC sample.

2.2.3. HPLC trace of 5mC, 5hmC, 5fC, and 5caC

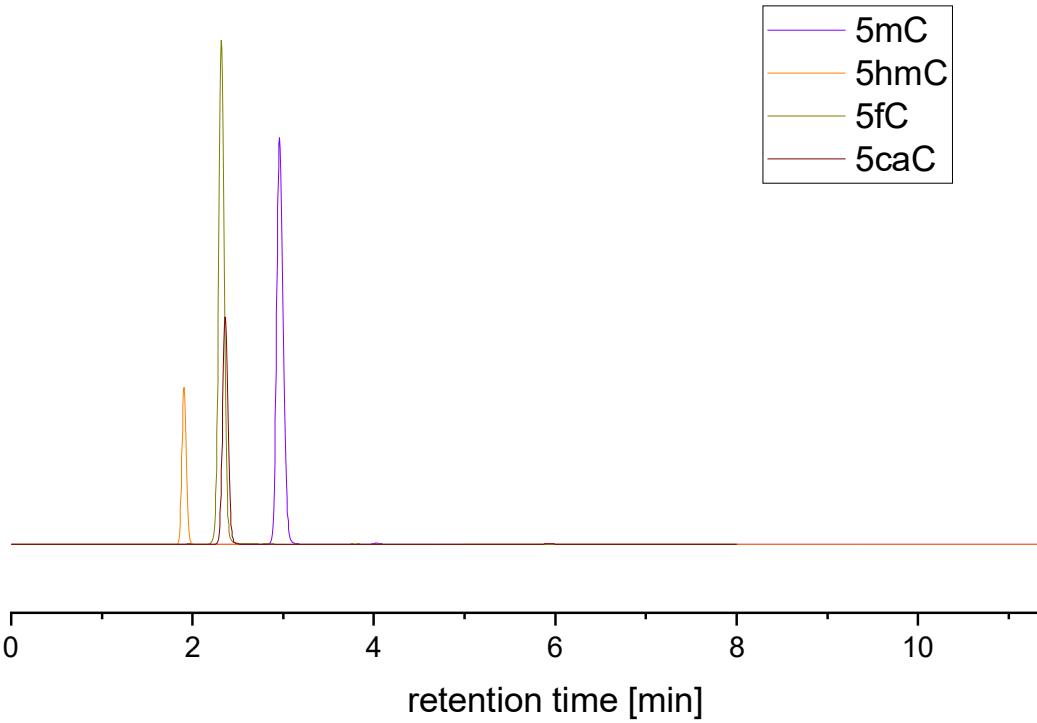


Figure S4. Superimposed HPLC references traces of 5mC, 5hmC, 5fC, 5caC.

2.2.4. HPLC trace of **1** after filtration

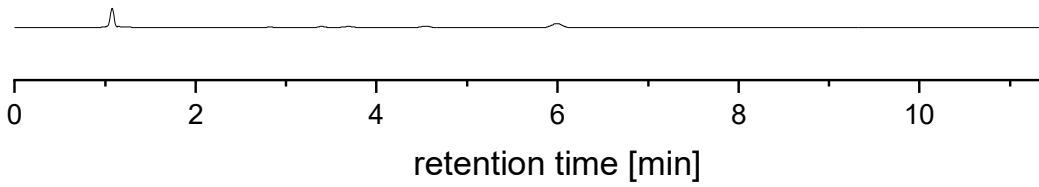


Figure S5. HPLC references trace of a sample of **1** that was treated like a regular reaction (filtration through silica, lyophilisation) as described under Section 2.1.

2.2.5. Spike Experiments – No Additions

In these experiments, the expected products were added individually to a product mixture obtained from the reaction of **1** with 5mdC ([**1**] = 5 mM, [5m
mixture analysed using HPLC. In the following, the original HPLC without any additions (**Figure S6**) as well as the spiked traces are shown. Thes
nucleoside/nucleobase pairs (**Figure S7** - **Figure S10**).

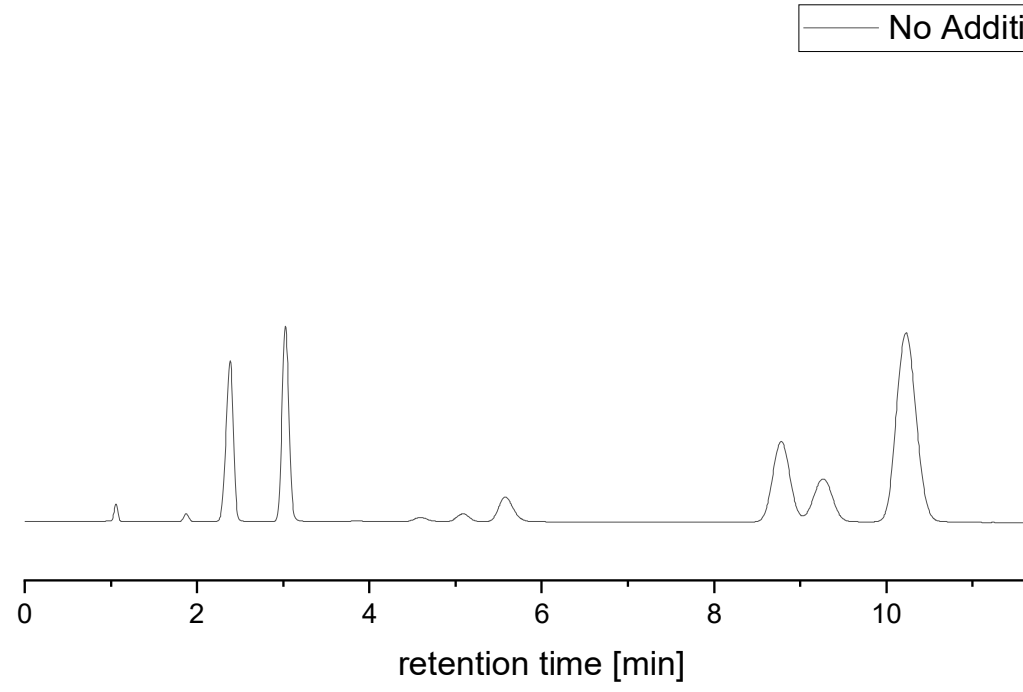


Figure S6. HPLC trace of a representative reaction sample of the reaction of **1** with 5mdC.

2.2.6. Spike Experiments – 5mdC or 5mC

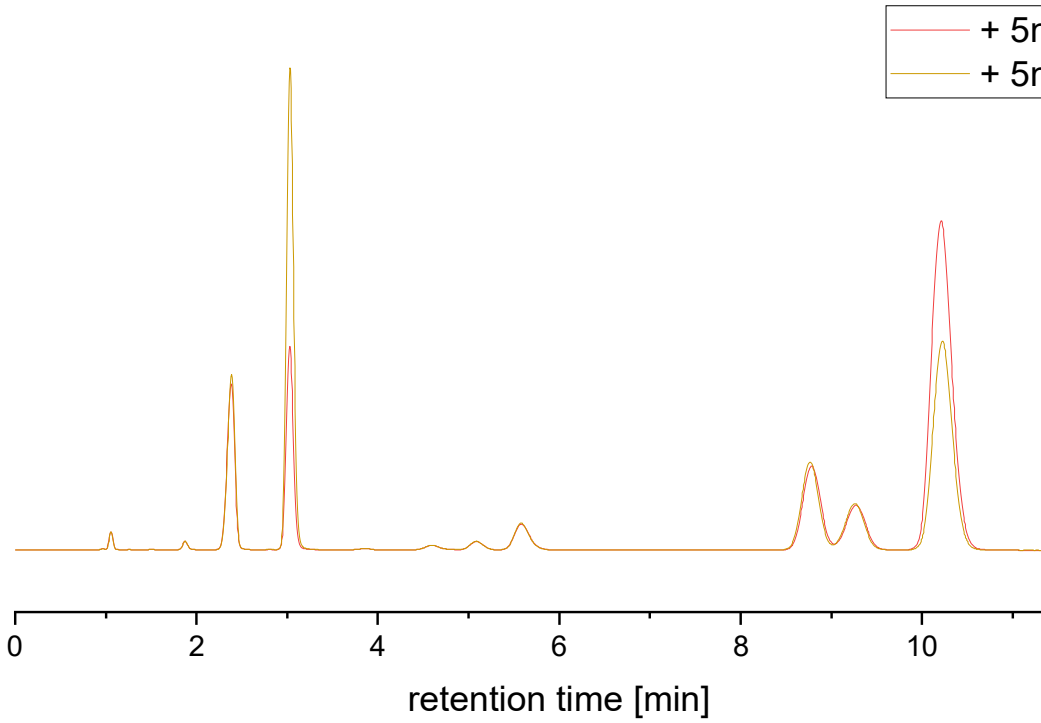


Figure S7. Superimposed HPLC traces of a representative reaction sample of the reaction of **1** with 5mdC with added 5mdC or 5mC.

2.2.7. Spike Experiments – 5hmdC or 5hmC

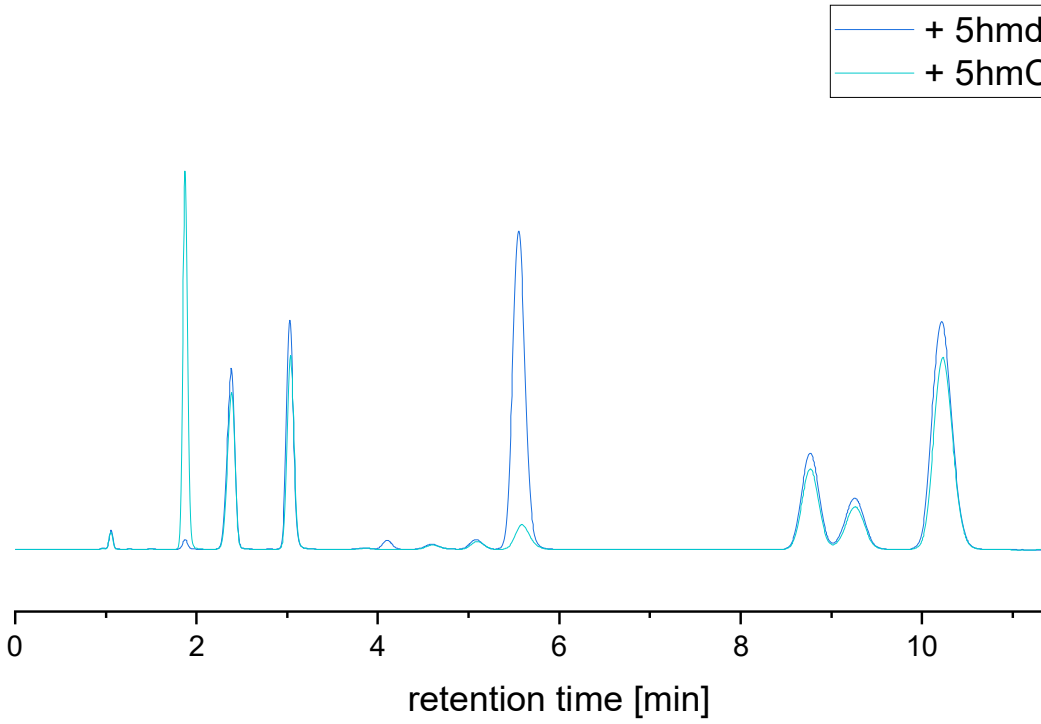


Figure S8. Superimposed HPLC traces of a representative reaction sample of the reaction of **1** with 5mdC with added 5hmdC or 5hmC.

2.2.8. Spike Experiments – 5fdC or 5fC

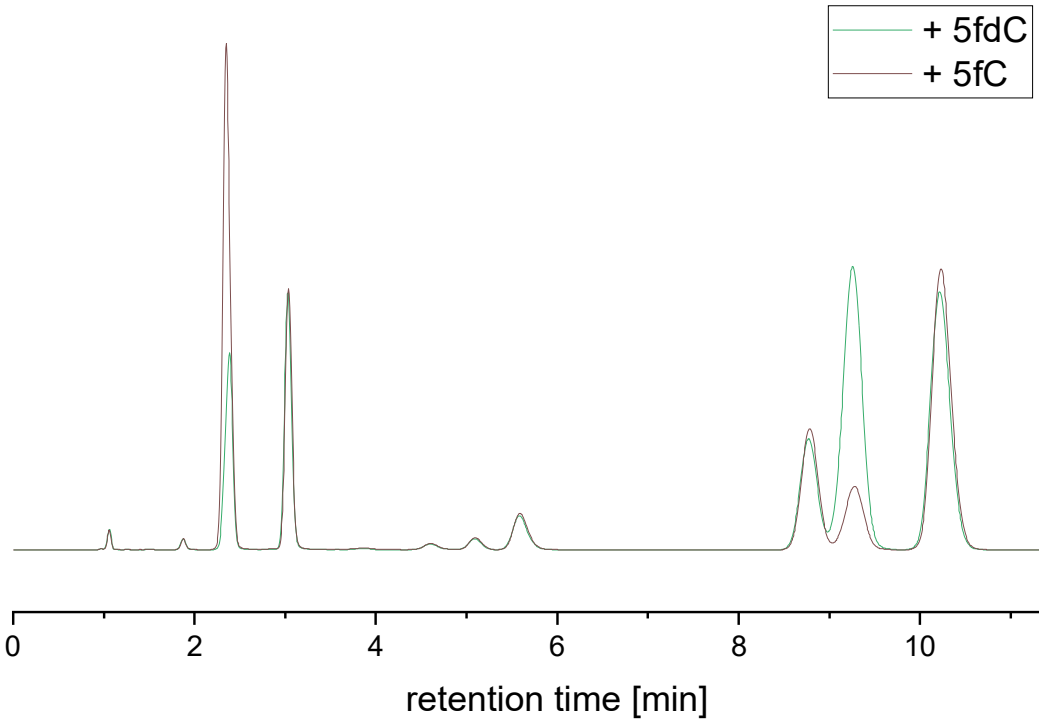


Figure S9. Superimposed HPLC traces of a representative reaction sample of the reaction of **1** with 5mdC with added 5fdC or 5fC.

2.2.9. Spike Experiments – 5cadC or 5caC

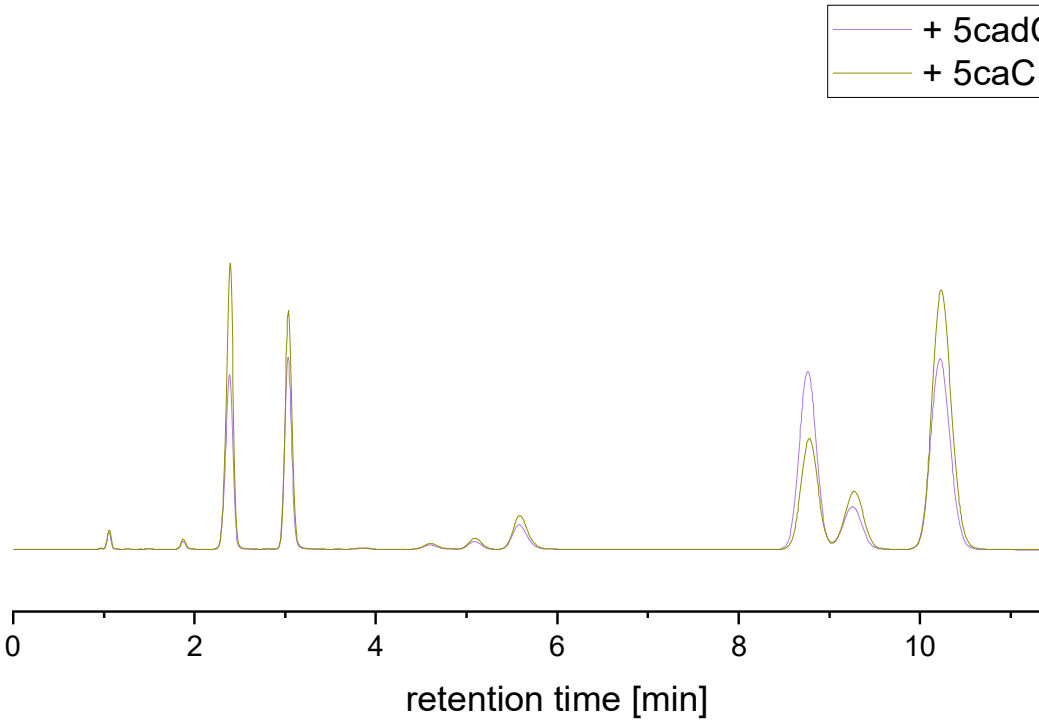


Figure S10. Superimposed HPLC traces of a representative reaction sample of the reaction of **1** with 5mdC with added 5cadC or 5caC.

2.3. GC-MS Reference Samples

2.3.1. 2'-Deoxyribono-γ-lactone

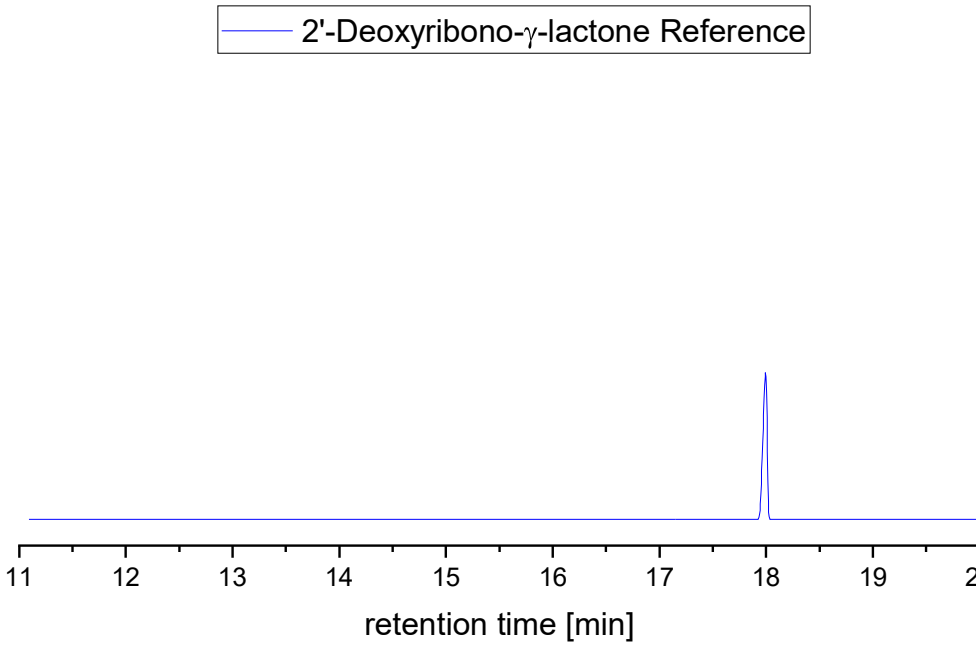


Figure S11. GC-MS trace of 2'-deoxy-ribono-γ-lactone as reference sample.

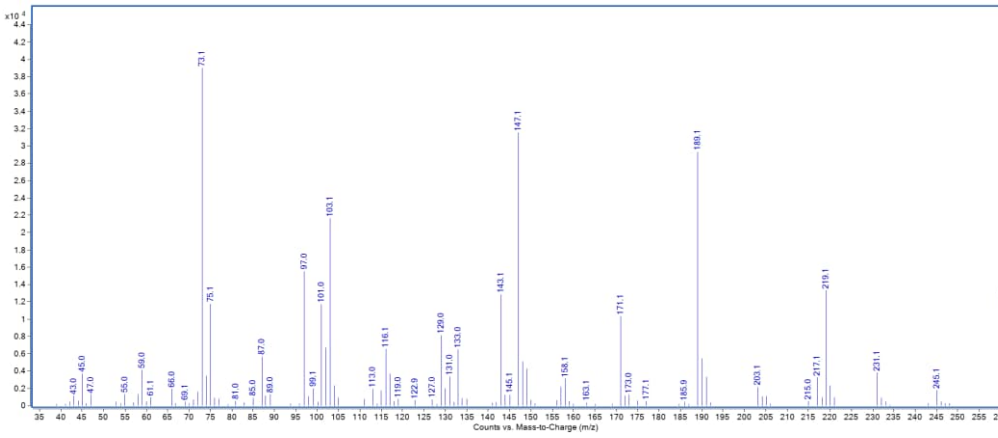


Figure S12. EI MS Spectrum at a retention time of 18.0 min, assigned to 2'-deoxyribono-γ-lactone, confirmed by 2020 NIST database.

2.3.2. 5-methyl-2'deoxyribo-cytidine

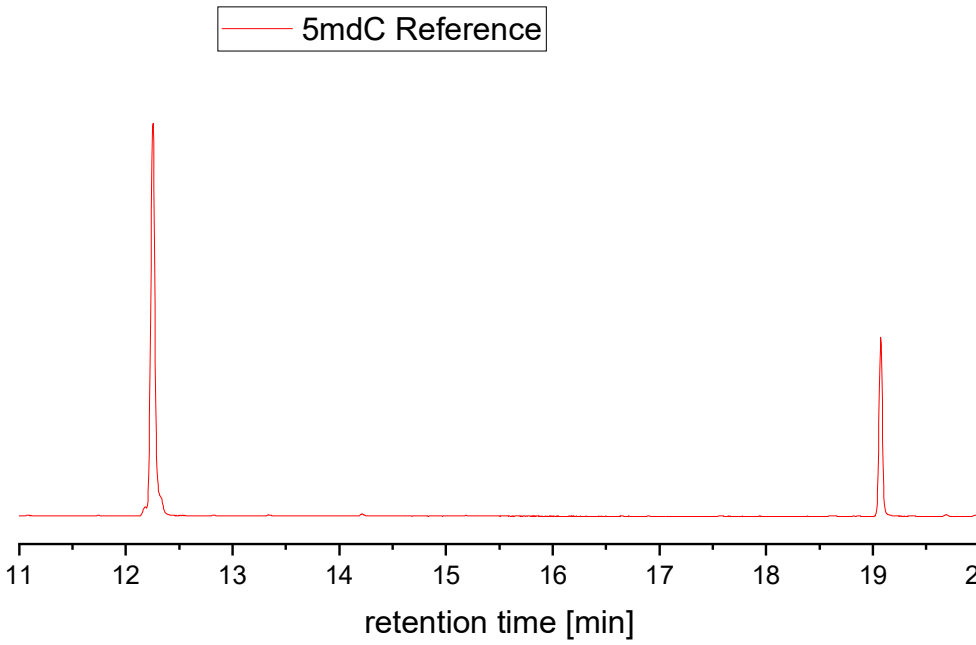


Figure S13. GC-MS trace of 5mdC as reference sample.

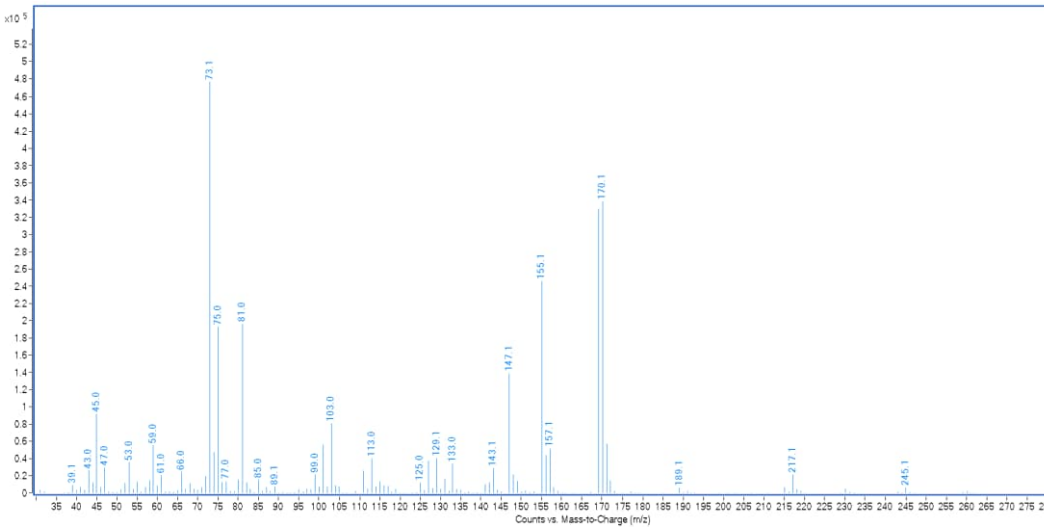


Figure S14. EI MS Spectrum at a retention time of 12.2 min, assigned to 2'-deoxyribose.

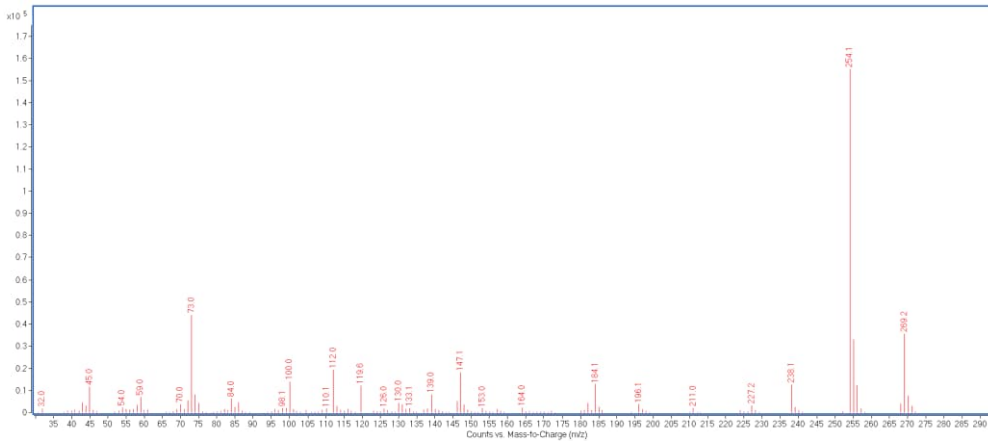


Figure S15. EI MS Spectrum at a retention time of 19.1 min, assigned to 5mC.[5]

2.3.3. Overlay of References

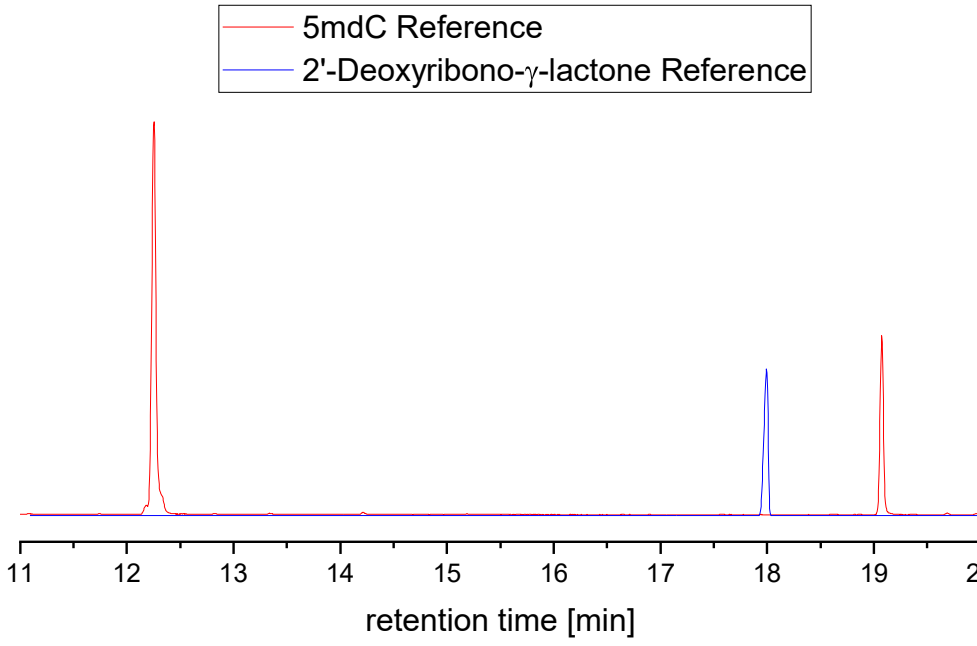
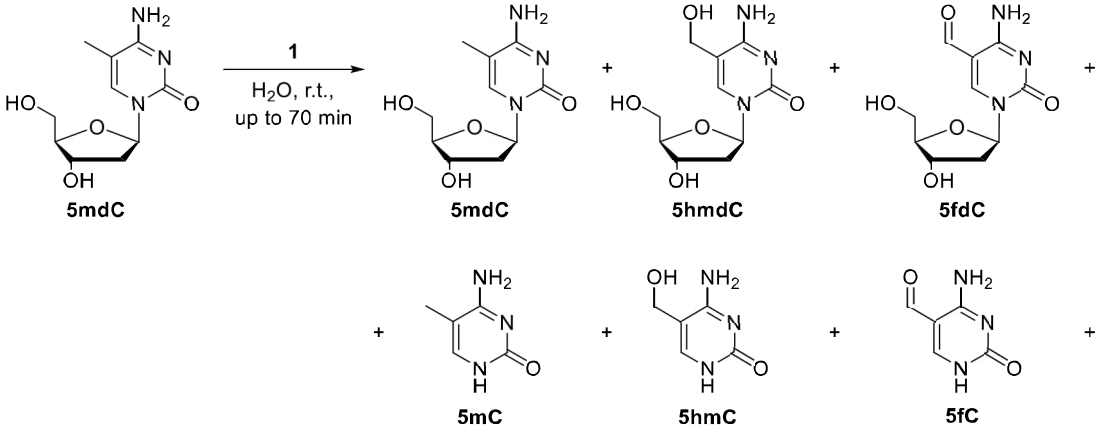


Figure S16. Superimposed GC-MS traces (compare **Figures S13** and **S14**) of 2'-deoxy-ribono-γ-lactone and 5mdC as reference samples.

2.4. Reactions of **1** with substrates

2.4.1. Reaction of **1** with 5mdC



Scheme S1: Reaction of 5mdC with 1 in water and observed products. Conditions: [5mdC] = 1 mM; [**1**] = 5 mM; T = 22 °C; t = 15 min or 30 min (re
up to 70 min (time-resolved reaction).

Reagent ratio experiments

To a solution of 5mdC in water (2 mM, 2.5 mM, 3.3 mM, 5 mM, 10 mM; 500 µL each) was added the appropriately prepared solution of **1** in wa
was performed twice. One series was then stirred for 15 min, the other one for 30 min during which a color change from green to slightly brow
the reactions were filtered through silica as described above. The freeze-dried residues were resuspended in water (either 1000 mL for 1:1
subjected to HPLC analysis (5.0 µL injection) as described above.

Time-resolved reaction

To a solution of 5mdC in water (2.0 mM, 1.60 mL, 1.0 equiv.) was added the appropriately prepared solution of **1** in water (10 mM, 1.60 mL,
during which a color change from green to orange-brown was observable. Every 5 min, samples (200 µL) were taken as described above. T
water (200 µL) and subjected to HPLC analysis (5.0 µL injection) as described above.

Representative HPLC trace (1:5 equivalents, 30 min)

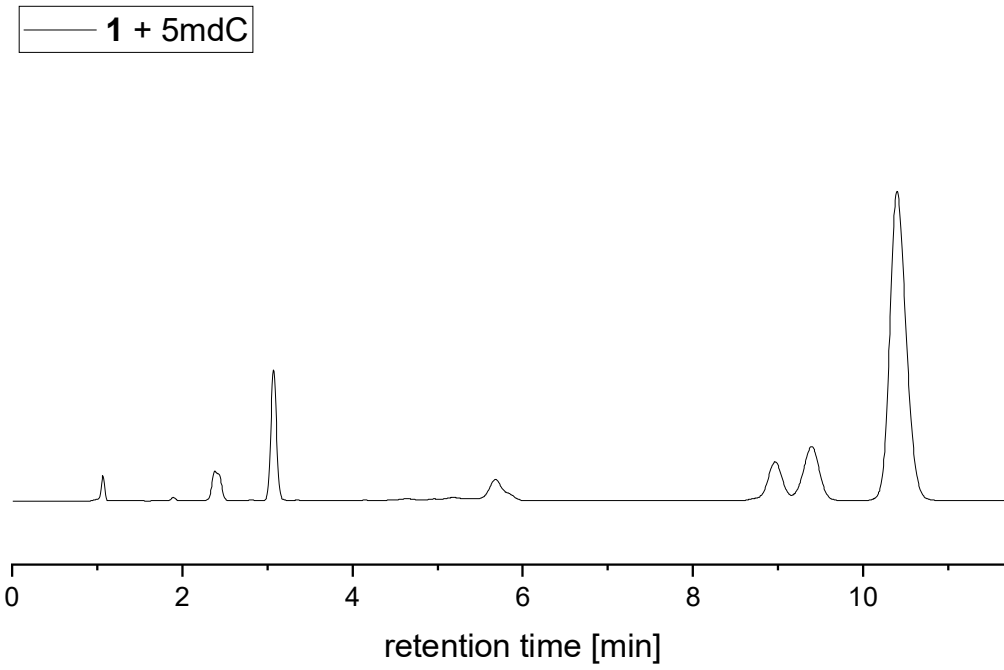


Figure S17. HPLC trace of a representative reaction sample of the reaction of **1** (5 mM) with 5mdC (1 mM) after 30 min.

Mass spectrometric analysis of the reaction

A reaction between 5mdC and **1** (1.00 mL, final concentrations: [5mdC] = 1.0 mM, [**1**] = 5.0 mM) was carried out for 70 min and worked up a
was utilized for LC-MS and GC-MS measurements.

For LC-MS, the reaction sample was resuspended in water to the original 5mdC starting concentration (max. 1.0 mM). The experiment was
used for HPLC experiments and with 100% aqueous NH_4HCO_2 buffer pH 5.5.

For GC-MS the reaction sample was suspended in acetonitrile (200 uL) and BSTFA (200 uL) added. The mixtures were heated to 70 °C for 3

GC-MS trace (1:5 equivalents, 30 min)

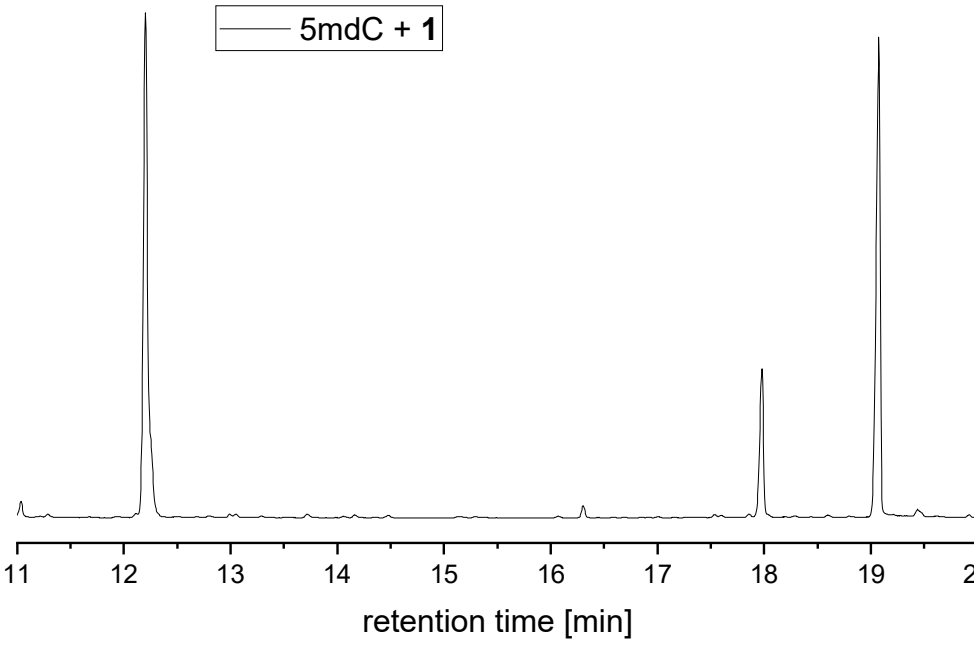


Figure S18. GC-MS trace of a representative reaction sample of the reaction of **1** (5 mM) with 5mdC (1 mM) after 30 min.

Overlaid with reference samples

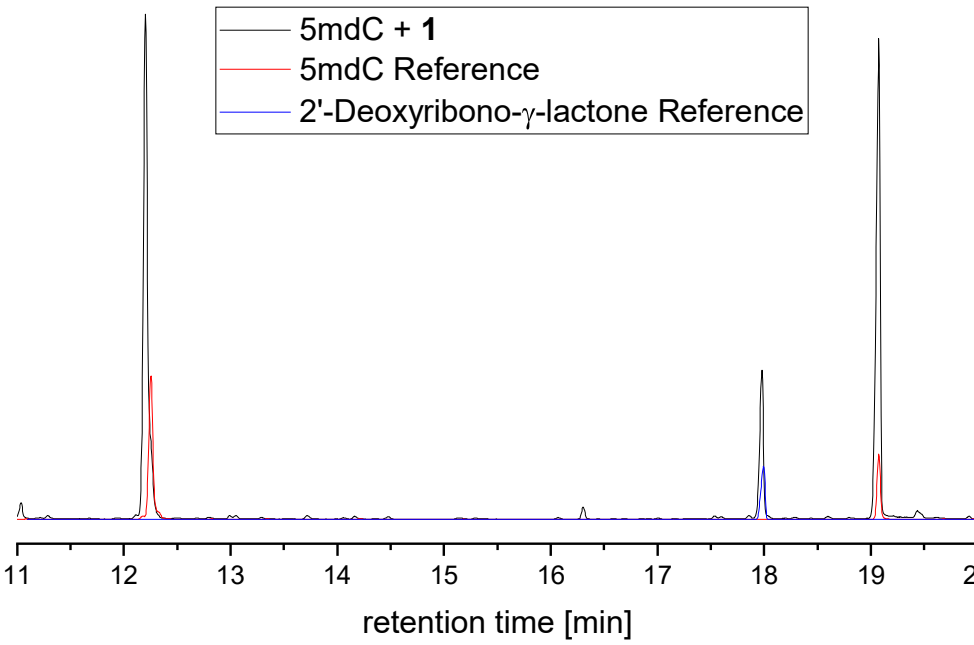


Figure S19. GC-MS trace of a representative reaction sample of the reaction of **1** (5 mM) with 5mdC (1 mM) after 30 min overlaid with references.

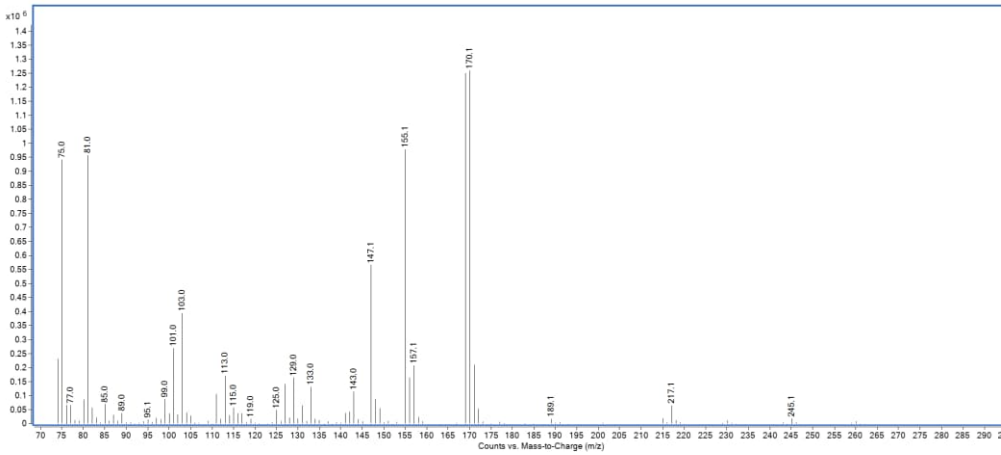


Figure S20. EI MS Spectrum at a retention time of 19.1 min, assigned to 2'-deoxyribose.

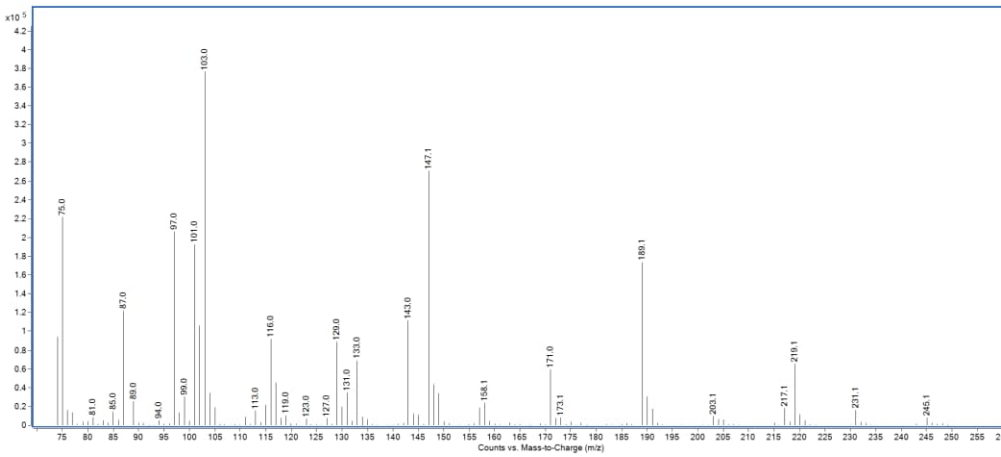


Figure S21. EI MS Spectrum at a retention time of 19.1 min, assigned to 2'-deoxyribono-γ-lactone, confirmed by 2020 NIST database.

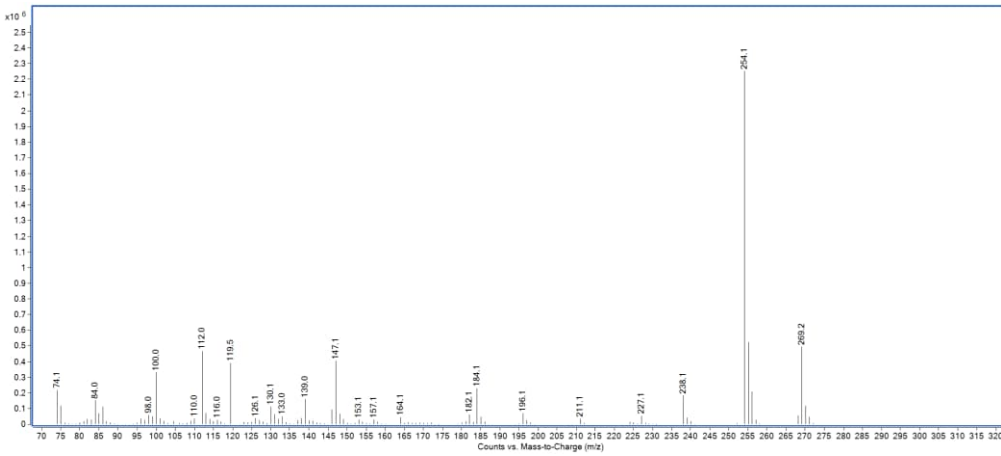
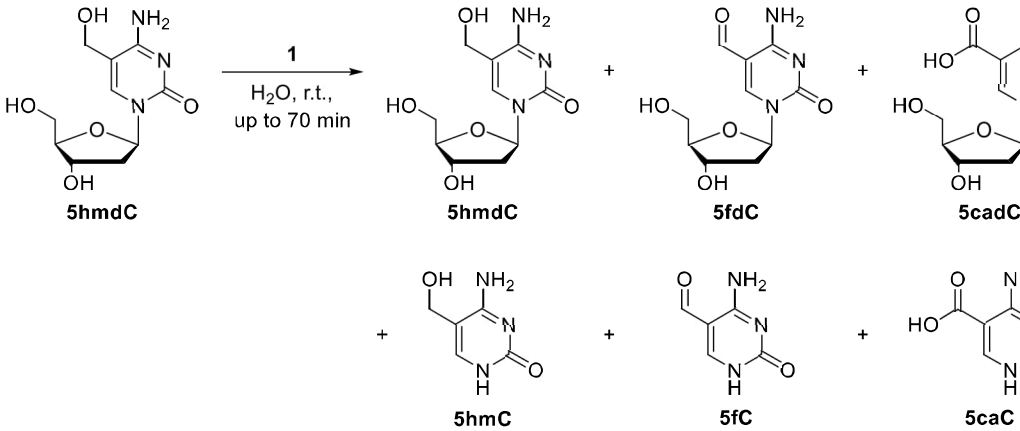


Figure S22. EI MS Spectrum at a retention time of 19.1 min, assigned to 5mC[5], confirmed by 2020 NIST database.

2.4.2. Reaction of **1** with 5hmdC



Scheme S2: Reaction of 5hmdC with **1** in water and observed products. Conditions: [5hmdC] = 1 mM; [**1**] = 5 mM; T = 22 °C; t = up to 70 mi

This reaction was performed identical to the reaction with 5mdC as substrate, as described in Section 2.1. The observed color chang
5mdC.

Representative HPLC trace (1:5 equivalents, 30 min)

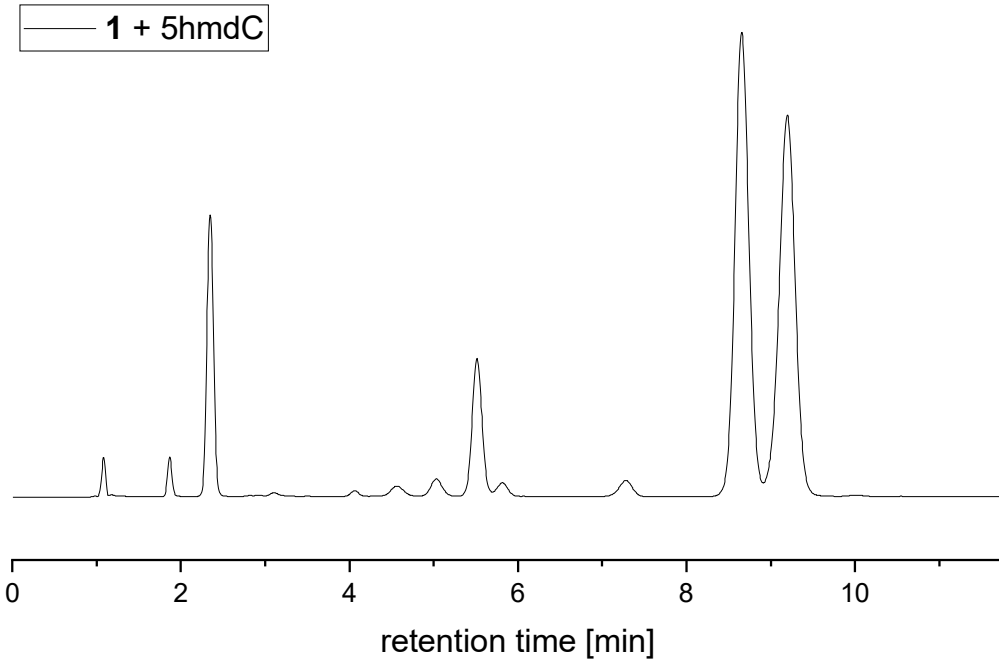


Figure S23. HPLC trace of a representative reaction sample of the reaction of **1** (5 mM) with 5hmdC (1 mM) after 30 min.

Nucleoside quantification

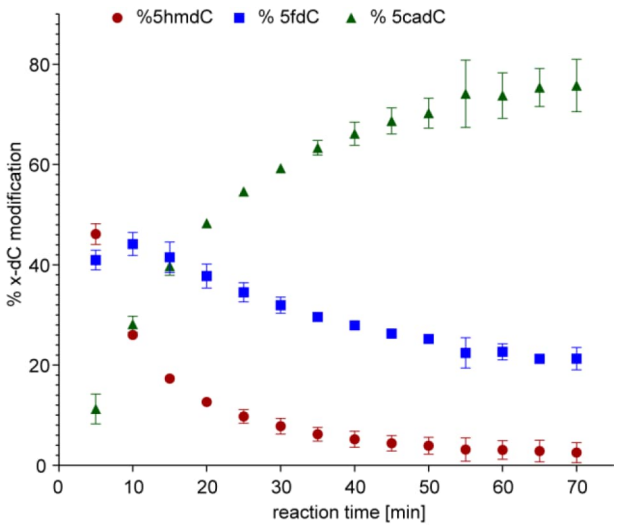
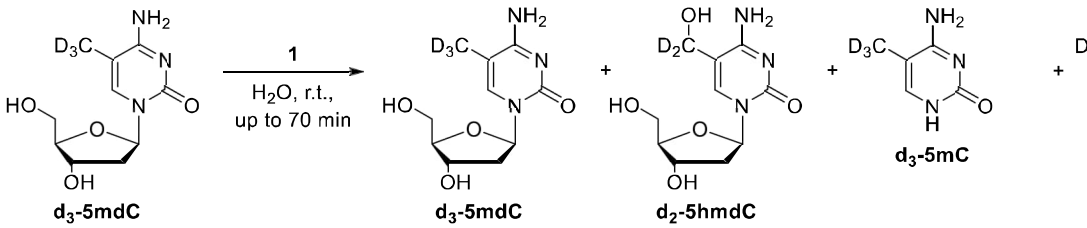


Figure S24. Detected amounts of 5hmdC, 5fdC, and 5cadC in the reaction between **1** and 5hmdC after certain time intervals (duplicate measurement). Conditio
points are averaged from two replicates.

2.4.3. Reaction of **1** with d_3-5mdC



Scheme S3: Reaction of D_3-5mdC with 1 in water and observed products. Conditions: [5hmdC] = 1 mM; [**1**] = 5 mM; T = 22 °C; t = up to 7

This reaction was performed identical to the reaction with 5mdC as substrate, as described in Section 2.1. The observed color chang
5mdC.

Representative HPLC trace (1:5 equivalents, 30 min)

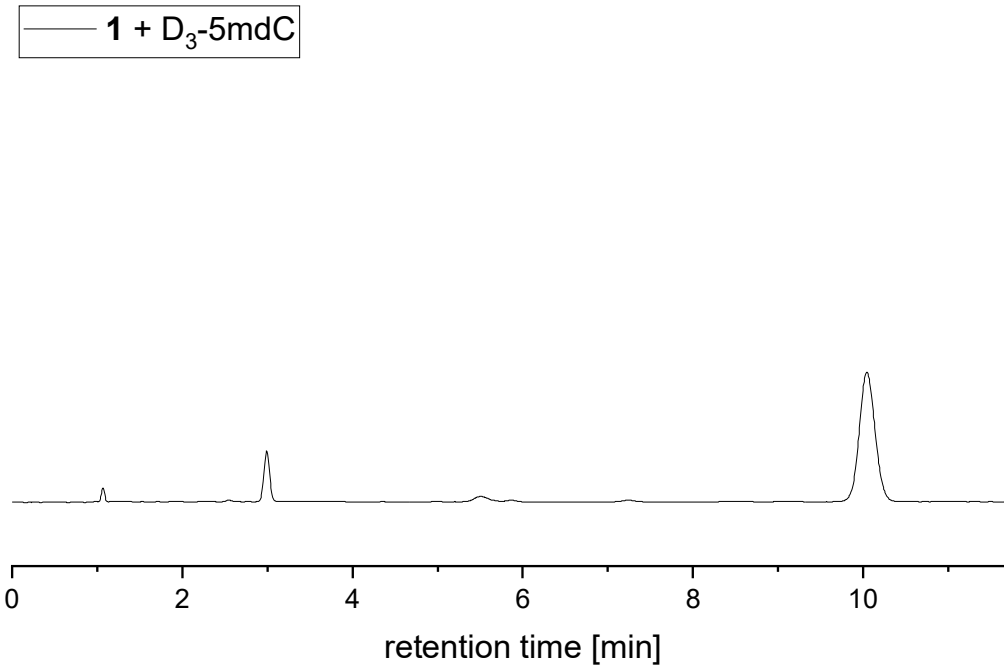


Figure S25. HPLC trace of a representative reaction sample of the reaction of **1** (5 mM) with D_3-5mdC (1 mM) after 30 min.

Nucleoside/Nucleobase Quantification

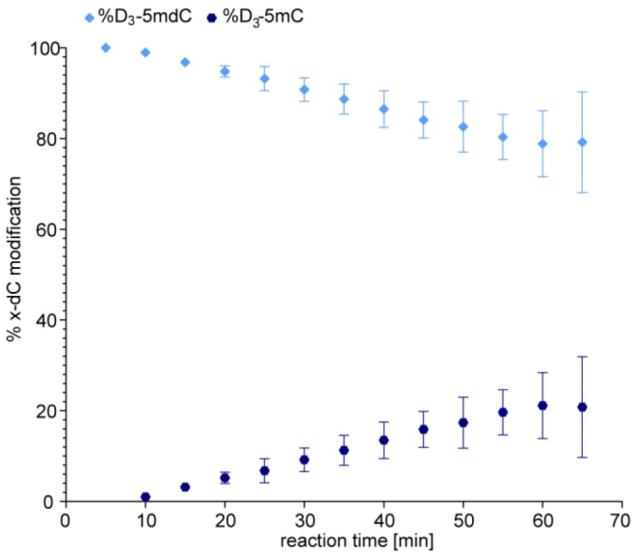


Figure S26. Detected amounts of D_3-5mdC and D_3-5mC in the reaction between **1** and D_3-5mdC after certain time intervals. Conditions: [D_3-5mdC] = 1 mM, [**1**] = two replicates.

2.4.4. Filtration and control experiments

To test whether all products were completely eluted (to ensure accuracy of quantification) under the filtration conditions described above mixtures of the eight monitored reaction products (either 0.02 mM, 0.20 mM or 1.0 mM) was filtered through silica (2.4 mL). A second identica dried residues were resuspended in water (200 µL) and the supernatant subjected to HPLC analysis (5.0 µL injection) as described above. Similarly, it was tested whether the nucleoside products would hydrolyze during the work-up procedure. A solution of all four nucleoside p through silica (2.4 mL) and worked up as described above. The freeze-dried residue was resuspended in water (200 µL) and the supernatant

3. Oligonucleotide Experiments

3.1. General Procedure for Oligonucleotide Experiments

The oligonucleotides were dissolved in MilliQ water so that a concentration of c = 0.5 mM or c = 1 mM was reached. The required amount of reaction tube and diluted with water to reach the desired concentration (see individual experiments for details). The iron(IV)-oxo complex [F described by Jonasson and Daumann to remove the cerium counter ion to obtain **1** with a mixture of F^-, OH^-, and NO_3^- counter ions and dilu Then, the required amount of **1** was added to the Eppendorff reaction tubes containing the oligo sample so that a final volume of v = 50 µl wa at room temperature (22-25 °C) for t = 45 min. Subsequently, the samples were diluted with another 50 µl of water and filtered through a pad and Daumann (0.3 ml silica in a 1 ml syringe, washed with 2 x 400 µl water).[5] The samples were lyophilized to dryness and then re-suspen (PTFE, 45 µm) the oligonucleotide concentration of the samples was determined *via* UV/Vis absorption at λ = 260 nm (see chapter Materials For time resolved reaction monitoring, a sample **O** ([**O**] = 0.25 mM) in 250 µl water was prepared. An aqueous solution of **1** (250 µl, [**1**] = 1 m of [**O**] = 0.125 mM and [**1**] = 0.5 mM was reached. The mixture was incubated at room temperate (22 °C) and samples (50 µl) were taken a 17 min, 26 min, 37 min and treated as described above. The samples were then treated as described above. The obtained solid product r water.

3.2. Oligonucleotide Concentration

The concentration of the oligonucleotide was determined as described above (see Materials and Methods). Two measurements of A_{260} were ta From this, the concentration was calculated ($c_{calc.}$) using Equation **1**. Then, the background generated during the experimental procedure (se which was then used for calculation of the necessary amounts for the digestion procedure.

Table S6: Overview of the conducted oligonucleotide experiments, the respective measured absorption values at λ = 260 nm and the thereby calculated concentrations. F coefficient ε_{260} supplied in **Table S4** was used.

Oligo 2							
Experiment	**Experimental Conditions**			**A_{260}**		**Concentration**	
	c_{Oligo} [mM]	**C_{Fe} [mM]**	**Meas. 1**	**Meas. 2**	**Avg.**	**$c_{calc.}$ [µM]**	**c_{corr} [µM]**
Untreated	0.50	0.00	1.034	1.059	1.047	11.8	10.4
O-2 + Fe1	0.25	0.25	4.017	4.073	4.045	45.6	41.5
O-2 + Fe2	0.25	1.00	4.146	4.176	4.161	46.9	42.8
TD1-0.5 min	0.125	0.5	0.910	1.010	0.960	10.7	9.3
TD1-5 min	0.125	0.5	1.192	1.366	1.279	14.2	12.8
TD1-10 min	0.125	0.5	1.145	1.064	1.105	12.3	10.9
TD1-17 min	0.125	0.5	1.396	1.219	1.308	14.5	13.2
TD1-26 min	0.125	0.5	1.050	1.107	1.079	12.0	10.6
TD1-37 min	0.125	0.5	1.108	1.120	1.114	12.4	11.0
TD2-0.5 min	0.125	0.5	1.091	1.115	1.103	12.3	10.9
TD2-2.5 min	0.125	0.5	1.306	1.606	1.456	16.2	14.8
TD2-5 min	0.125	0.5	1.092	1.184	1.138	12.6	11.3
TD2-10 min	0.125	0.5	1.150	1.228	1.189	13.2	11.8
TD2-17 min	0.125	0.5	1.248	1.368	1.308	14.5	13.2
TD2-26 min	0.125	0.5	1.223	1.388	1.306	14.5	13.1
TD2-37 min	0.125	0.5	0.698	0.835	0.767	8.5	7.1
ds-O*	0.125	0.5	1.726	1.570	1.648	9.9	9.1
ss-O*	0.125	0.5	1.002	1.145	1.074	11.9	5.7
ss-C*	0.125	0.5	0.820	0.798	0.809	7.8	4.1
No5mC-C	0.25	1.0	1.693	1.460	1.577	18.1	14.0
No5mC-CP	0.25	1.0	2.612	2.520	2.566	29.5	28.1
O-2 control	0.25	0.00	4.489	4.487	4.488	50.5	46.5
Fe control	0.00	1.00	0.366	0.346	0.356	4.0	0
H_2O**	0.00	0.00	0.138	0.109	0.124	1.4	0

*) These reactions contained additionally 2.5 mM NaCl. **) This sample was treated as all reaction samples (incubated, filtered, lyophilized, determination of the background absorbance at λ = 260 nm.

3.3. Digestion and Quantification Procedure

0.1 µg of oligonucleotide in 35 µL H_2O were digested as follows: 1 µL of Degradase plus (*Zymo research*), 2.5 µL of Degradase plus reaction b nuclease (*Merck, Serratia marcescens*) and a spesific amount of isotopically labelled internal standars were added. The mixure was incubat Prior to LC-MS/MS analysis, samples were filtered by using an AcroPrep Advance 9 filter plate 0.2 µm Supor (*Pall Life Science*).

3.4. Quantification Results

Here we present some additional data on the detection of 8-oxo-dG in the reaction samples of **1** with **O** (**Figure S27**, lower graph). Also, individual nucleoside levels detected in the time-dependent reaction of **1** with **O** (**Figure S28**). The data is the same as in Figure 4 in t presentation meant to provide more clarity.

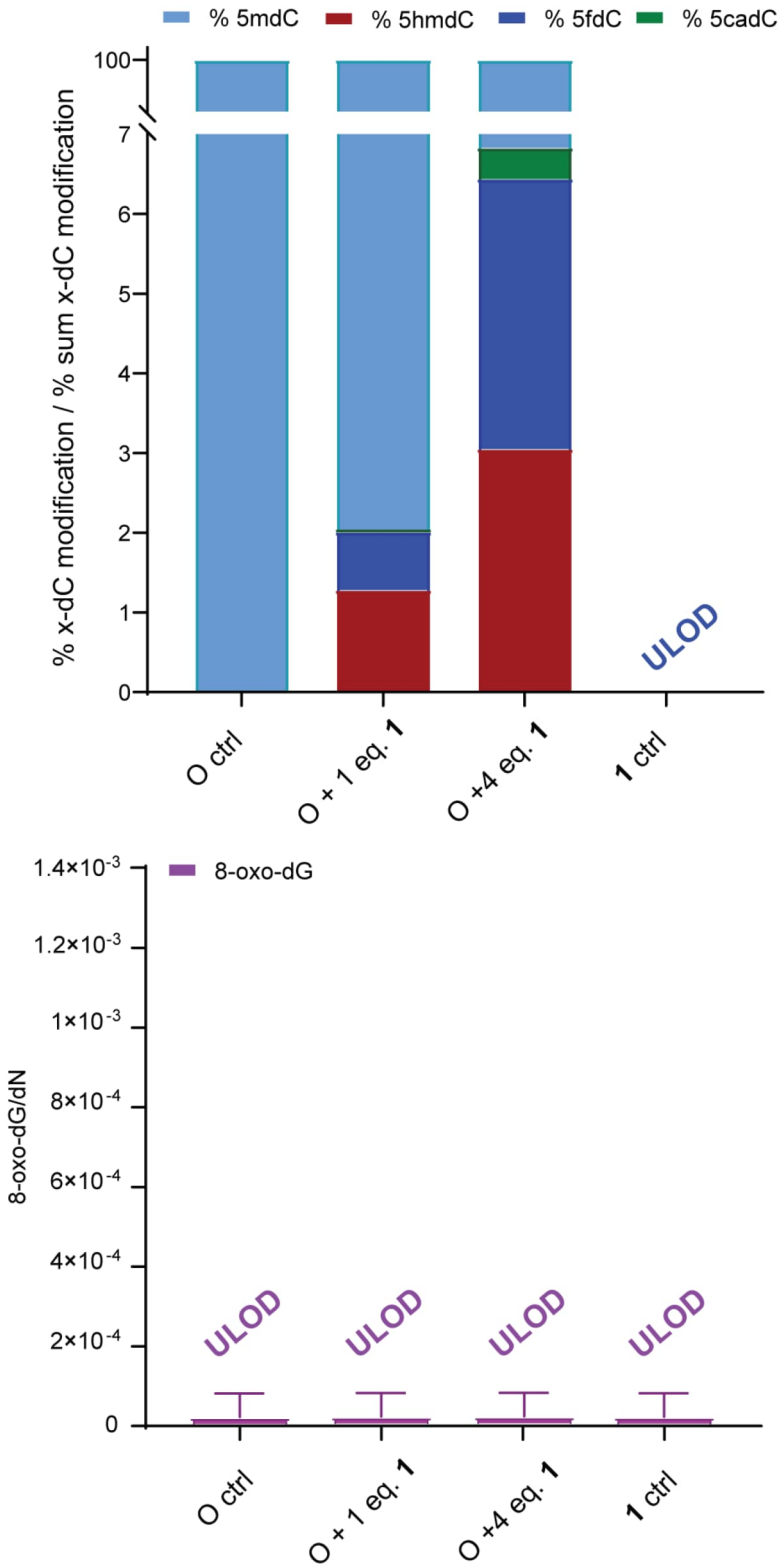


Figure S27: Detected amounts of 5mdC, 5hmdC, 5fdC, 5cadC (upper graph) and 8-oxo-dG (lower graph) in reaction of **1** with **O**. Conditions: [**O**] = 0.25 mM, [**1**] = 0.25 mM ULOD refers to under level of detection.

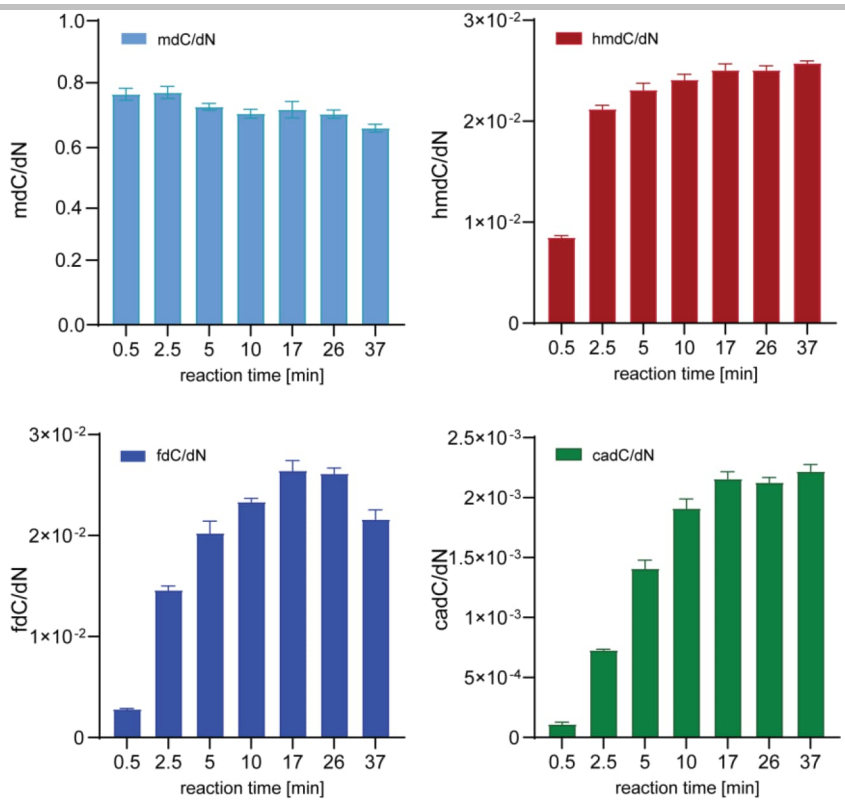


Figure S28: Right: Detected amounts of 5mdC (upper left), 5hmdC (upper right), 5fdC (lower left), and 5cadC (lower right) in reaction of **1** with **O** over time. Conditions: [

3.5. Product Ratios

In the reaction of **1** with the oligonucleotide **O**, we detected different product ratios depending on the amount of **1** that was used. **Figure S29** equiv. of **1**, respectively.

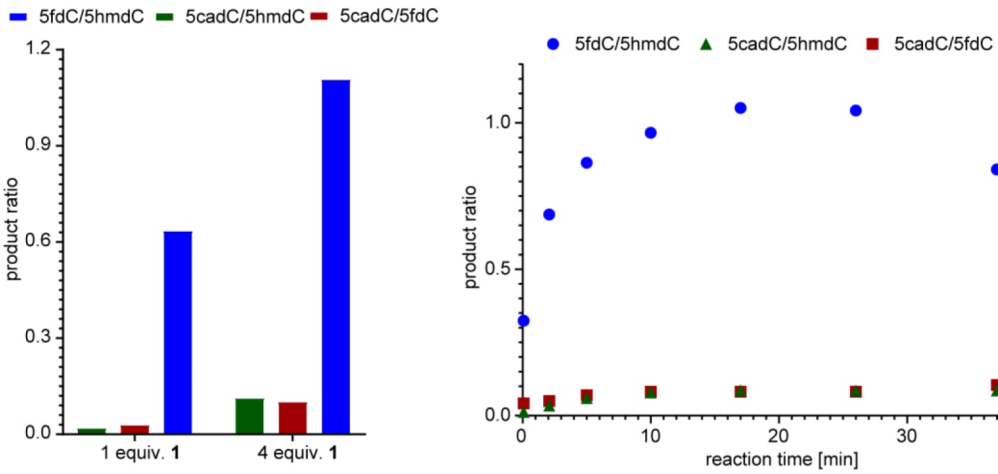


Figure S29: Left: Product ratios 5cadC/5hmdC, 5cadC/5fdC, and 5fdC/5cadC for different equivalents of **1** (1 or 4 equiv.). **Right:** Product ratios 5cadC/5hmdC, 5cadC/5f to 37 min).

As expected, a higher amount of **1** leads to more oxidized species. It remains to be tested whether the continuous addition of **1** leads to acc Whereas the ratios 5cadC/5hmdC and 5cad/5fdC quickly approach a constant value (0.08-0.1) the ratio between 5fdC/5hmdC shows a ma This could prove useful to stir the reaction towards producing mostly 5fdC formation.

3.6. Side-Product Analysis

In the following, characteristic MALDI mass spectra are discussed in detail and possible fragmentation pathways are proposed.

O control

A signal at ~2620 m/z was observed in all measurements, including the **O** control (**Figure S30**, treated as reaction mixture without expos probably stems from production and was attributed to an 9-mer oligonucleotide. Additionally, a signal at ~3205 m/z was also observed i oligonucleotide sample. This signal probably also stems from production and was attributed to an 11-mer oligonucleotide. Very small amou also observed but could not be assigned to any species.

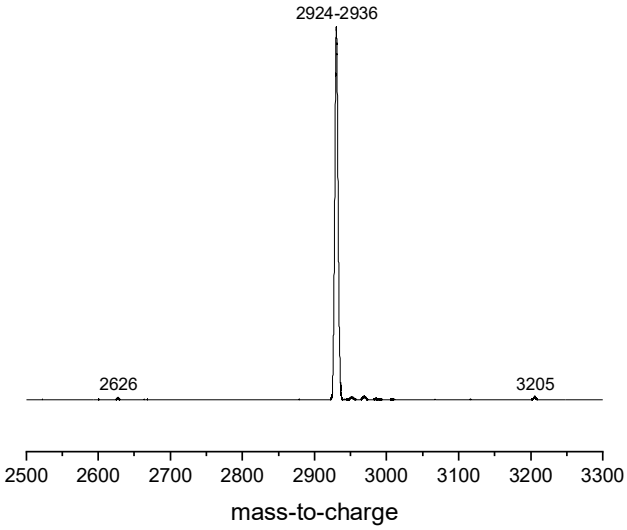


Figure S30: Excerpt of the MALDI MS spectrum obtained for the control reaction of O without **1**. Small signals are observed at 2626 m/z and 3205 m/z, these were assign production.

O + 1 equiv. **1**

In the reaction of **O** with 1 equiv. of **1** several additional signals compared to the control sample were observed. In **Figure S31** a possibl additional amounts of the signal at 2626 m/z is shown. During this pathway, a cytidine-3'-phosphate residue is generated. It is unclear whethe only residual amounts of the 9-mer (vide supra) are detected. In **Figure S32**, two oxidative pathways are shown resulting in the loss of gu signals at 2810 m/z and 2834 m/z, respectively. These proposed pathways are further discussed in **Scheme S4**.

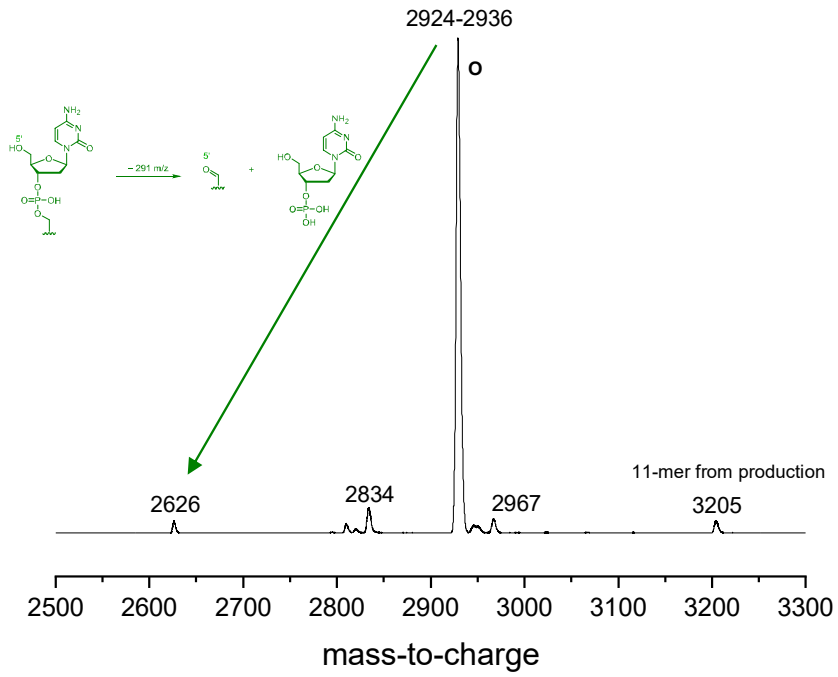


Figure S31: Excerpt of the MALDI MS spectrum obtained for the reaction of **O** with 1 equiv. of 1. A possible fragmentation pathway resulting in the observed signal at 262 observed in all samples and attributed to an 11-mer oligonucleotide contamination from production of the oligonucleotide. The signals between 2800-3000 m/z are shown

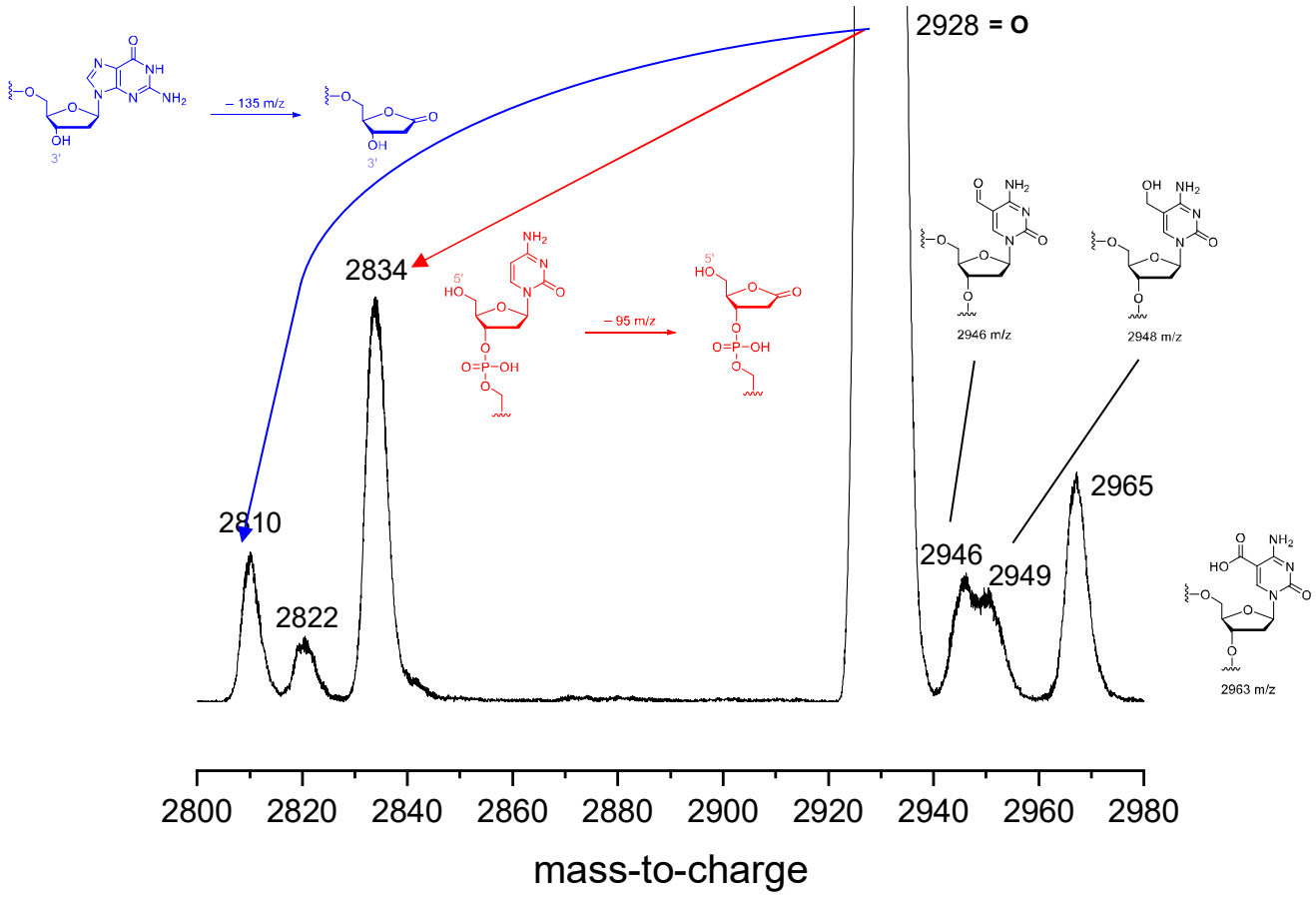


Figure S32: Zoomed in excerpt of the MALDI MS spectrum obtained for the reaction of **O** with 1 equiv. of **1**. Possible fragmentation pathway resulting in the observed sig guanine or cytosine, blue and red, respectively). The signal at 2946, 2949 and 2965 m/z were attributed to 5hmdC, 5fdC, and 5cadC-containing 10-mer oligonucleotides s 2822 m/z could not be assigned clearly to any side reaction, however, loss of adenine might present a possible explanation.

O + 4 equiv. **1**

In this sample, the same fragmentation pattern is detected in MALDI MS as in the one that was only exposed to 1 equiv. of **1** (**Figure S31** an reacted/decomposed oligonucleotides are observed. Additional small signals were detected that could not be assigned.

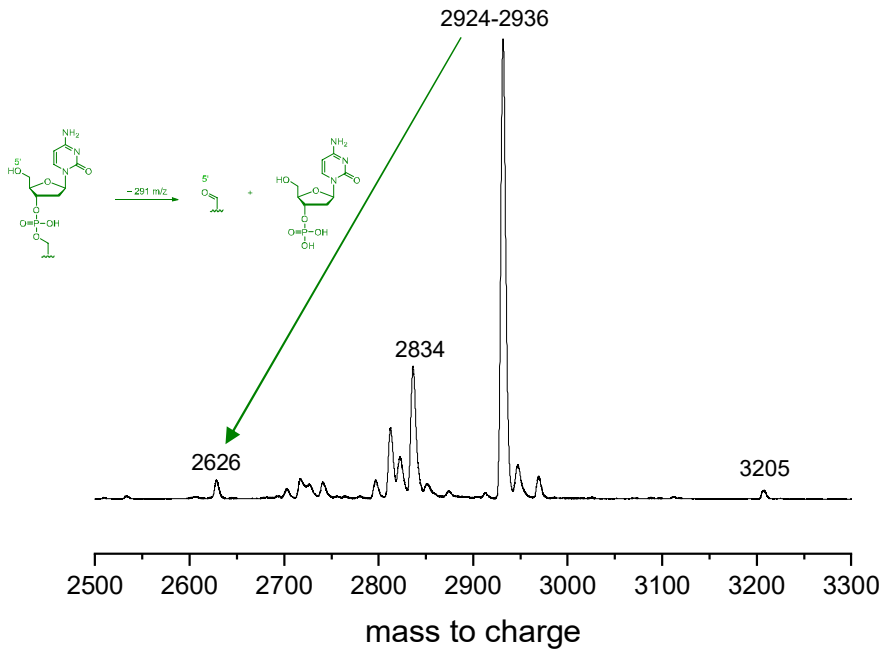


Figure S33: Excerpt of the MALDI MS spectrum obtained for the reaction of **O** with 4 equiv. of 1. A possible fragmentation pathway resulting in the observed signal at 262 3000 m/z are shown enlarged in **Figure S34**, those between 2680-2940 m/z are shown enlarged in **Figure S35**.

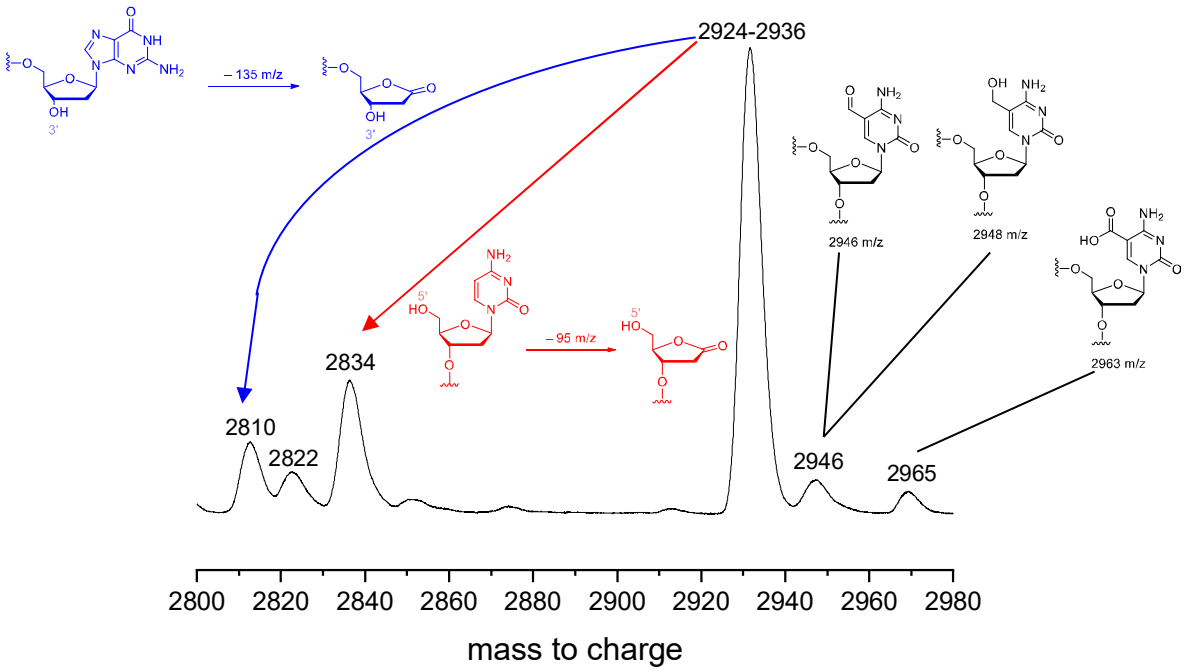


Figure S34: Zoomed in excerpt of the MALDI MS spectrum obtained for the reaction of **O** with 4 equiv. of **1**. Possible fragmentation pathway resulting in loss of guanine o

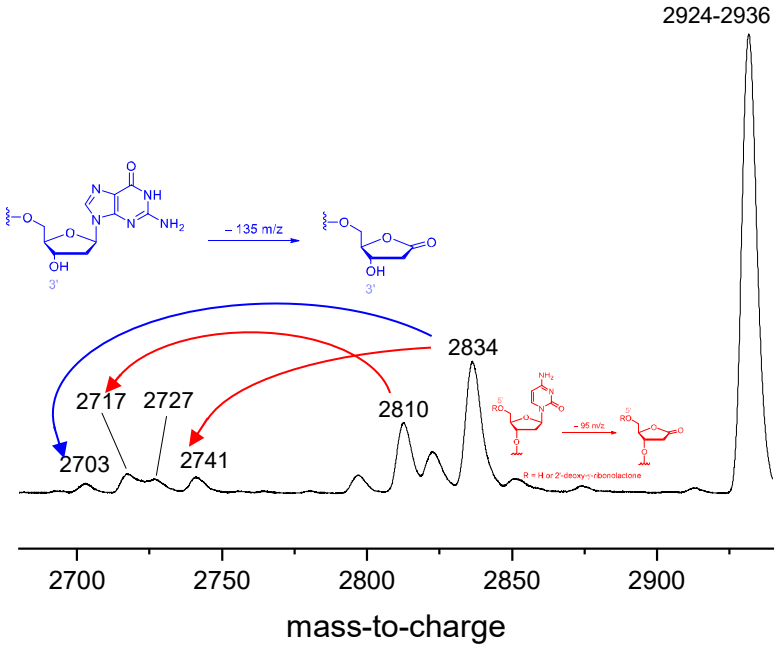


Figure S35: Zoomed in excerpt of the MALDI MS spectrum obtained for the reaction of O with 4 equiv. of **1**. Possible fragmentation pathway resulting in loss of an additio and red: 2834 m/z to 2741 m/z or 2810 m/z to 2717 m/z) are indicated. Fragments with m/z at 2834 and 2810 resulted from a previous loss of a cytosine or guanidine frag

Control Reactions: **O-2** and **O-2P**

In order to screen for additional side reactions, we exposed the "control strand" **O-2** HO-CCUUAACCCG-OH to 4 equiv. of **1**. Using MALDI M detected a very significant amount of decomposition reactions – much more than if the regular oligonucleotide **O** is used. Loss of guanine ar material and from the decomposition product that resulted from loss of cytosine or cytidine-3'-phosphate in the first place.
We also wanted to see if the presence of a phosphate cap on the 5' end of the oligonucleotide had any effect on the decomposition reaction CCUUAACCCG-OH to 4 equiv. of **1**. We did the same fragmentation patterns as already discussed, however, it seems that less loss of cyto CCUUAACCCG-OH (**Figure S37**).

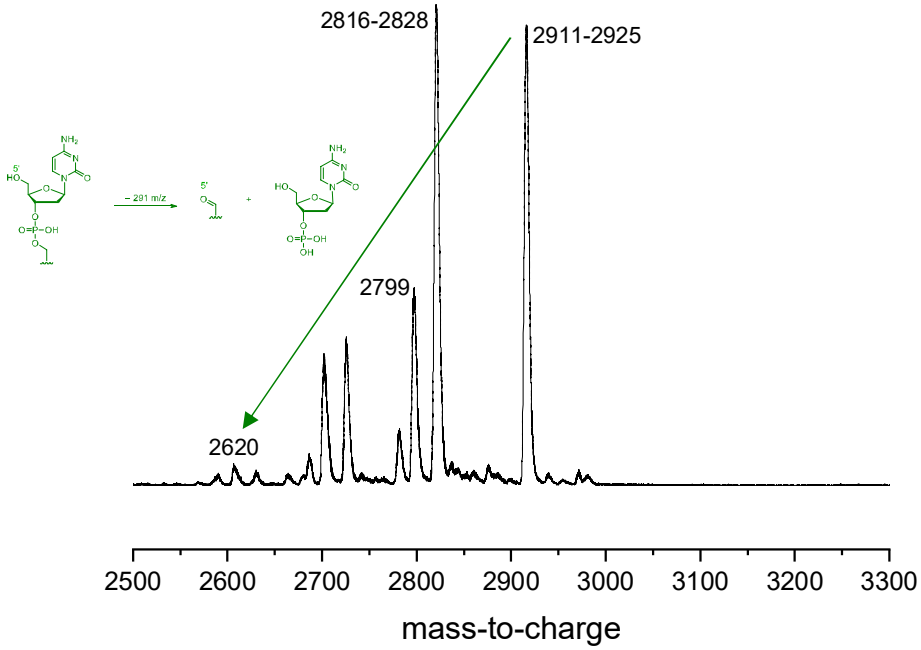


Figure S36: Excerpt of the MALDI MS spectrum obtained for the reaction of the "control strand" **O-2** HO-CCUUAACCCG-OH with 4 equiv. of **1**. A possible fragmentation is indicated (green). The signals between 2650-2950 m/z are shown enlarged in **Figure S37**.

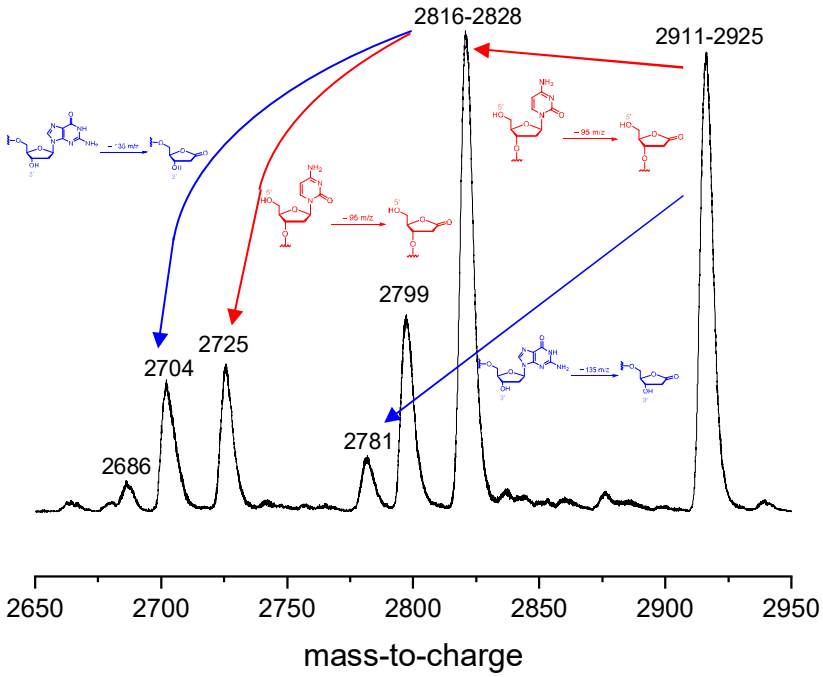


Figure S37: Zoomed in excerpt of the MALDI MS spectrum obtained for the reaction of the "control strand" **O-2** HO-CCUUAACCCG-OH with 4 equiv. of **1**. Possible fragm cytosine (blue and red, respectively) are indicated.

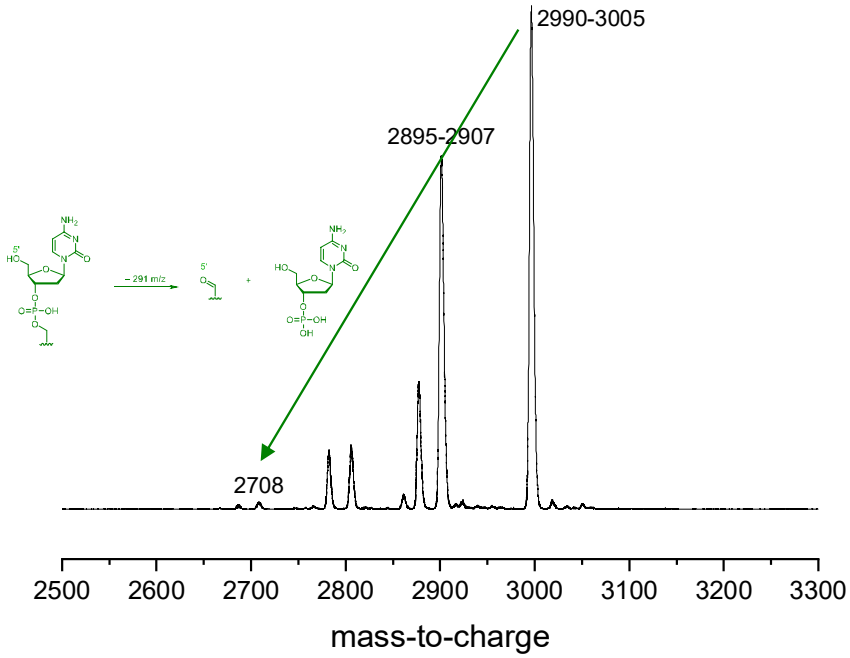


Figure S38: Excerpt of the MALDI MS spectrum obtained for the reaction of the "control strand" **O-2P** HO_3PO-CCUUAACCCG-OH with 4 equiv. of **1**. A possible fragment m/z is indicated (green). The signals between 2750-3050 m/z are shown enlarged in **Figure S37**.

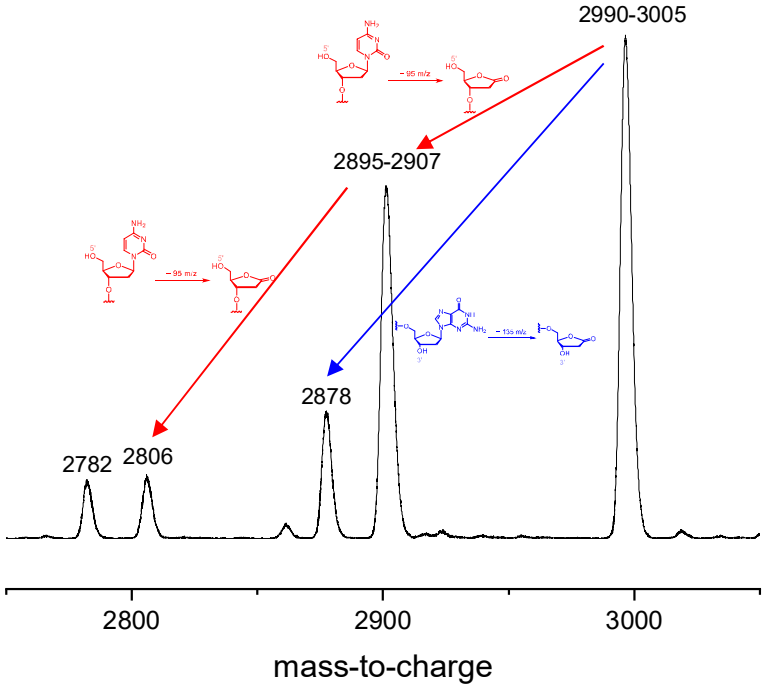
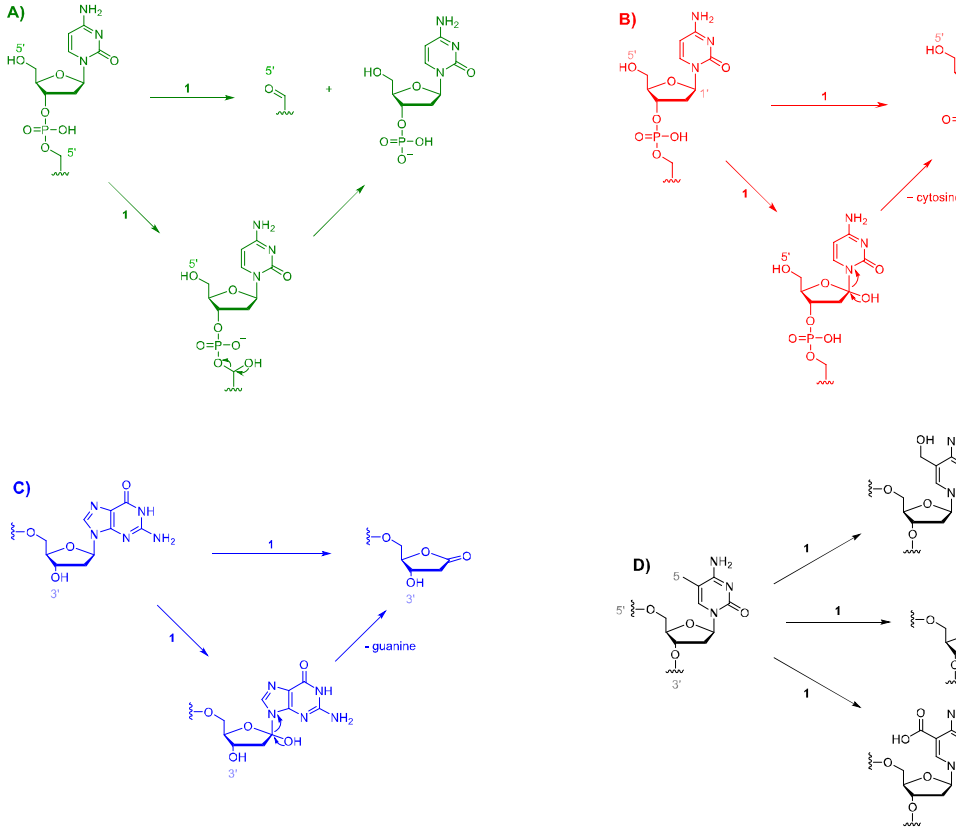


Figure S39: Zoomed in excerpt of the MALDI MS spectrum obtained for the reaction of the "control strand" **O-2P** HO_3PO-CCUUAACCCG-OH with 4 equiv. of **1**. Possible cytosine (blue and red, respectively) are indicated. The signal at 2782 m/z is also probably formed by loss of guanine from the signal at 2895-2907 m/z (not indicated by a

Mechanistic Proposal: Hydroxylation of Position 5' or 1':

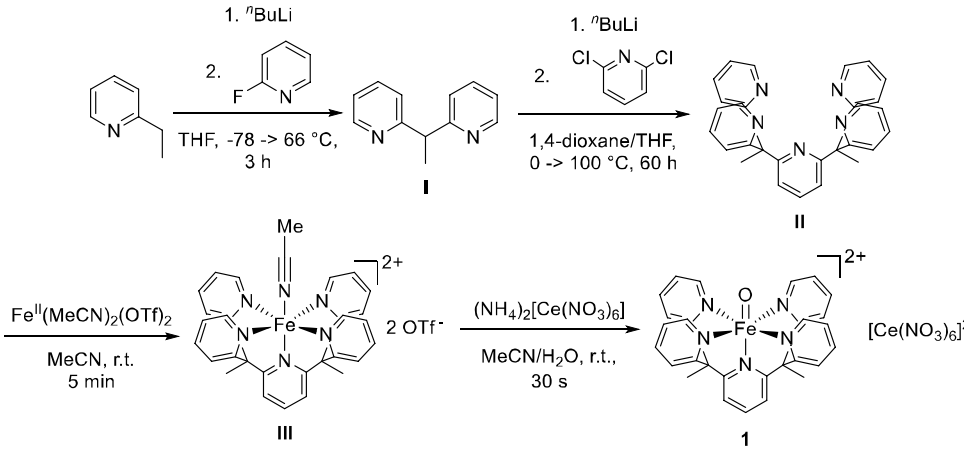


Scheme S4: A) Hydroxylation/oxidation of the 5' position of the second-to-last nucleotide on the 5' end of the oligonucleotide strand resulting in loss of a cytidine-3'-phos
on the nucleotide (green). **B)** Hydroxylation/oxidation of the 1' position of the last nucleotide on the 5' end of the strand resulting in loss of a cytosine nucleobase
Hydroxylation/oxidation of the 1' position of the last nucleotide on the 3' end of the strand resulting in loss of a guanine nucleobase fragment and formation of a lactone (b
on the second-to last position on the 3' end of the strand resulting the formation of 5hmdC, 5fdC, and 5cadC residues

4. Synthetic Procedures

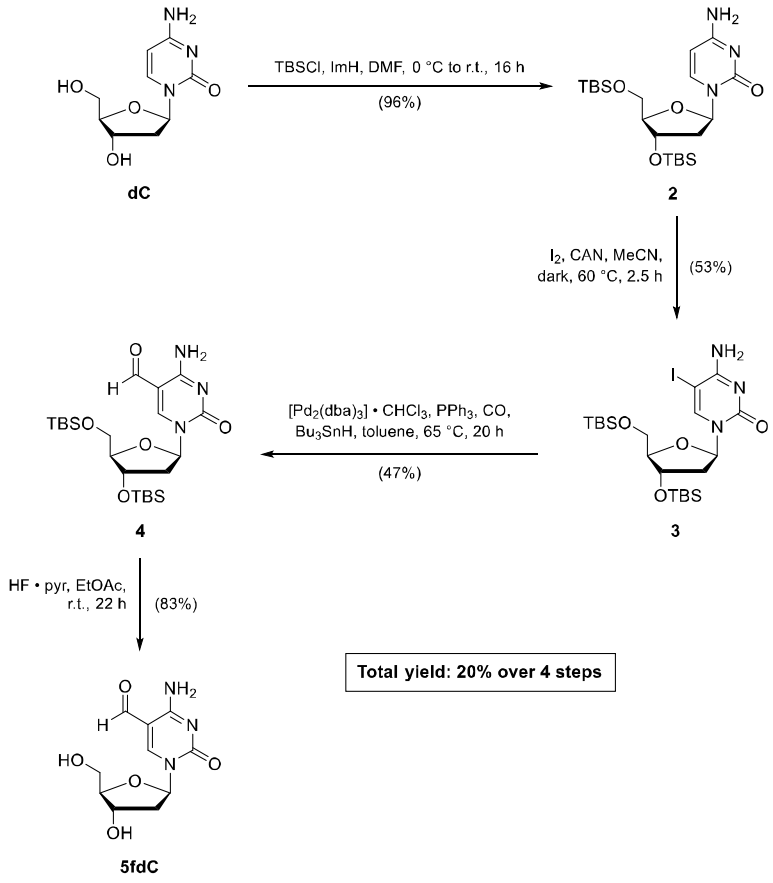
4.1. Synthesis of **1**

1[Ce(NO$_3$)$_6$)] was synthesized according to published literature procedures.[6,7] Exchange of the cerium counterion to a mixture of F$^-$, OH$^-$ a literature procedure.[5]



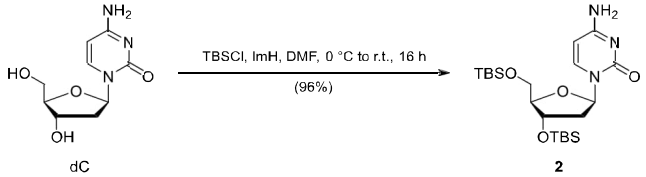
Scheme S5: Four step synthesis of [Fe(O)(Py$_5$Me$_2$-H)][Ce(NO$_3$)$_6$] as performed in this work, adapted from Cong *et al*.[7] and Chang *et al*.[6] A detailed discussion of the syn counterion can be found in our previous publication involving **1**.[5]

4.2. Synthesis of 5-formyl-2'deoxy-cytidine (5fdC)



Scheme S6: Four step synthesis of 5fdC as performed in this work, adapted from the literature.[8–10]

4.2.1. Synthesis of silyl protected 2'-deoxy-cytidine (**2**)



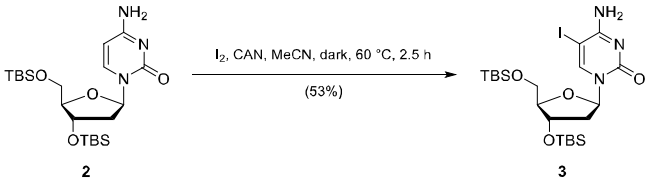
dC **2**

According to a modified standard literature procedure:[8]
Both 2'-deoxycytidine (dC; 1.00 g, 4.40 mmol, 1.00 equiv.) and imidazole (1.50 g, 22.0 mmol, 5.00 equiv.) were dissolved in anhydrous *N*,*N*- was cooled to 0 °C. *tert*-Butyldimethylsilyl chloride (1.99 g, 13.2 mmol, 3.00 equiv.) was added in three portions and the reaction was allowe 16 h, the mixture was diluted with ethyl acetate (30 mL) and saturated aqueous sodium bicarbonate solution (80 mL) was added slowly. The was extracted with ethyl acetate (3×50 mL). The combined organic layers were washed with saturated aqueous sodium chloride solution (50 solution was filtered and the filtrate was concentrated *in vacuo*. The resulting yellow viscous oil was purified by flash column chromatography afford silyl-protected 2'-deoxy-cytidine **2** (1.93 g, 4.23 mmol, 96%) as a foaming colorless solid.

^{1}H NMR (400 MHz, $CDCl_3$, 295 K):
δ = 8.03 (dd, *J* = 7.2, 3.2 Hz, 1H), 7.78 (s, 0H), 7.12 (s, 1H), 6.26 (dd, *J* = 6.4, 5.0 Hz, 1H), 5.66 (d, *J* = 7.3 Hz, 1H), 4.36 (q, *J* = 5.8 Hz, 1H), 2.43 (dt, *J* = 12.5, 6.0 Hz, 1H), 2.17 – 2.02 (m, 1H), 0.92 (s, 9H), 0.87 (s, 9H), 0.10 (d, *J* = 2.0 Hz, 5H), 0.05 (s, 6H).

HR-MS (ESI):
calculated for $C_{21}H_{42}N_3O_4Si_2$ $[M+H]^+$: 456.2708
found: 456.2705.

4.2.2. Synthesis of silyl protected 5-iodo-2'-deoxy-cytidine (**3**)



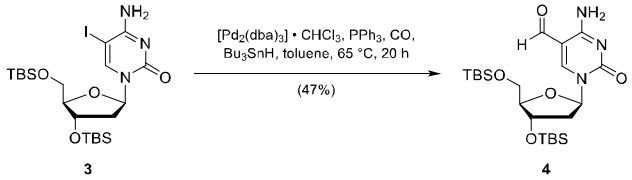
2 **3**

According to a modified standard literature procedure:[9]
Silyl-protected 2'-deoxycytidine **2** (1.79 g, 3.92 mmol, 1.00 equiv.) and iodine (2.19 g, 8.62 mmol, 2.20 equiv.) were dissolved in anhydrous a (4.73 g, 8.62 mmol, 2.20 equiv.) was added, upon which the solution turned dark brown, and the reaction was heated to 60 °C under exclu cooled to room temperature and saturated aqueous sodium bicarbonate solution (50 mL), followed by aqueous sodium thiosulfate solution yellow suspension. The suspension was diluted with water (100mL) and acetonitrile was evaporated under reduced pressure. The residual su and the filter cake was extensively washed with ethyl acetate (300 mL). The layers were separated and the aqueous layer was extracted organic layers were washed with saturated aqueous sodium chloride solution (100 mL) and dried over sodium sulfate. The dried solution wa *vacuo*. The resulting yellow solid was purified by flash column chromatography (0% to 2% to 4% methanol in dichloromethane) to afford sily 2.06 mmol, 53%) as a foaming colorless solid.

^{1}H NMR (400 MHz, $CDCl_3$, 295 K):
δ = 8.02 (s, 1H), 6.21 – 6.13 (m, 1H), 5.37 (s, 1H), 4.28 (dt, *J* = 6.0, 3.0 Hz, 1H), 3.91 (q, *J* = 2.7 Hz, 1H), 3.82 (dd, *J* = 11.4, 2.6 Hz, 1H), 3.6 13.3, 5.9, 3.1 Hz, 1H), 1.89 (ddd, *J* = 13.3, 7.1, 6.1 Hz, 1H), 0.84 (d, *J* = 19.8 Hz, 17H), 0.07 (d, *J* = 4.5 Hz, 6H).

HR-MS (ESI):
calculated for $C_{21}H_{41}IN_3O_4Si_2$ $[M+H]^+$: 582.1675
found: 582.1680.

4.2.3. Synthesis of silyl protected 5-formyl-2'-deoxy-cytidine (**4**)



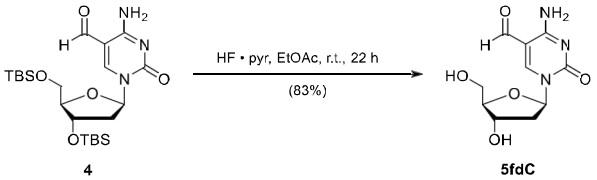
3 **4**

According to a modified standard literature procedure:[10]
Silyl-protected 5-iodo-2'-deoxycytidine **3** (600 mg, 1.03 mmol, 1.00 equiv.), [$Pd_2(dba)_3$] chloroform adduct (107 mg, 0.103 mmol, 0.100 eq
mmol, 0.600 equiv.) were dissolved in anhydrous toluene (15 mL) in a high-pressure reactor. The reactor was flushed with carbon mono
solution and the pressure was released after 10 min. This flushing procedure was repeated two more times to achieve carbon monoxide s
reactor was filled with carbon monoxide gas (3.5 bar) and the reaction mixture was heated to 65 °C. A solution of tributyltin hydride (0.33 mL
(0.33 mL) was added dropwise *via* syringe over 18 h. After 13 h of the addition, the pressure had dropped to normal pressure which is why t
gas (3.5 bar). After another 7 h, the reaction mixture was cooled to room temperature and the pressure was released. The greenish-black s
mL) and filtered through a plug of silica. The filter cake was extensively washed with ethyl acetate (200 mL) and the filtrate was concentrated *i*
by flash column chromatography on a mixture of silica gel and potassium carbonate (9:1; 20% to 33% to 50% ethyl acetate in hexanes) to a
4 (232 mg, 0.480 mmol, 47%) as a yellow solid.

^{1}H NMR (400 MHz, $CDCl_3$, 295 K):
δ = 9.49 (s, 1H), 8.56 (s, 1H), 8.19 (s, 1H), 6.21 (t, *J* = 6.1 Hz, 1H), 6.10 (s, 1H), 4.35 (dt, *J* = 6.4, 3.7 Hz, 1H), 4.05 (q, *J* = 2.7 Hz, 1H), 3.96
11.6, 2.5 Hz, 1H), 2.62 (ddd, *J* = 13.6, 6.2, 3.9 Hz, 1H), 2.14 – 2.02 (m, 1H), 0.90 (s, 9H), 0.89 (s, 8H), 0.10 (s, 3H), 0.09 (s, 3H), 0.08 (s, 3H

HR-MS (ESI):
calculated for $C_{22}H_{42}N_3O_5Si_2$ $[M+H]^+$: 484.2658
found: 484.2650.

4.2.4. Synthesis of 5-formyl-2'deoxyx-cytidine (5fdC)



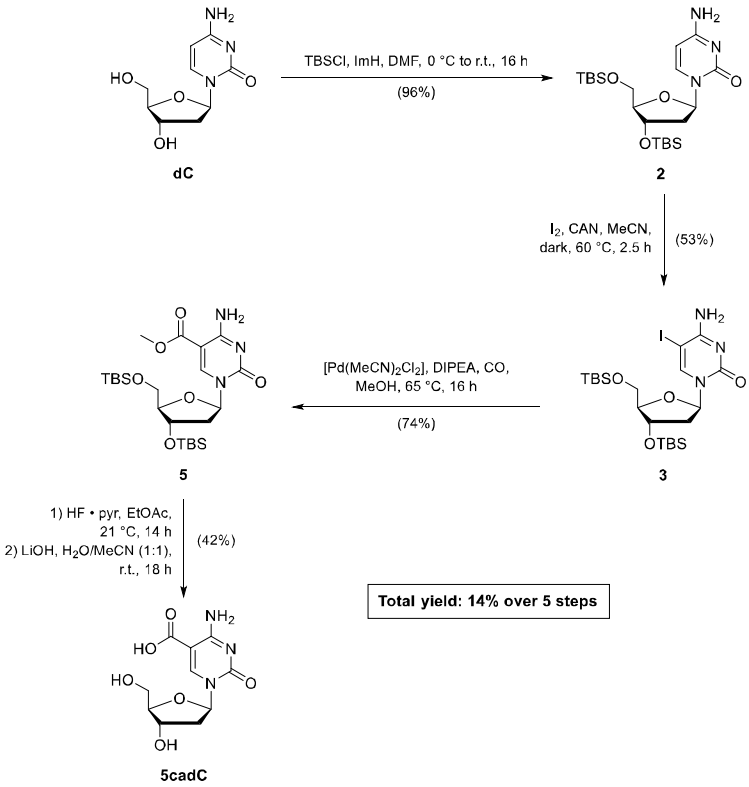
4 **5fdC**

According to a modified standard literature procedure:[10]
Silyl-protected 5-formyl-2'-deoxycytidine **4** (229 mg, 0.473 mmol, 1.00 equiv.) was dissolved in anhydrous ethyl acetate (14 mL) in a plastic tu
complex (70wt% HF, 0.37 mL, 30 equiv.) was added *via* syringe at room temperature. After 22 h, the yellow suspension was cooled to 0 °C a
64 equiv.) was added slowly *via* syringe. The reaction was allowed to slowly warm to room temperature and after additional two days, the re
mL) and all volatiles were removed *in vacuo*. The yellow, solid crude product was purified by HPLC (Column: Agilent Prep-C18 column 50×1
afford 5-formyl-2'-deoxycytidine (**5fdC**) (99.8 mg, 0.391 mmol, 83%) as a colorless solid.

^{1}H NMR (400 MHz, D_2O, 295 K):
δ = 9.53 (s, 1H), 8.82 (s, 1H), 6.21 (t, *J* = 6.1 Hz, 1H), 4.44 (dt, *J* = 6.6, 4.7 Hz, 1H), 4.15 (td, *J* = 4.7, 3.3 Hz, 1H), 3.92 (dd, *J* = 12.6, 3.3 Hz,
(ddd, *J* = 14.2, 6.5, 4.9 Hz, 1H), 2.38 (ddd, *J* = 14.2, 6.5, 5.6 Hz, 1H) ppm.

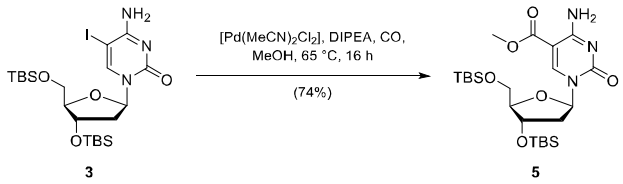
HR-MS (ESI):
calculated for $C_{10}H_{14}N_3O_5$ $[M+H]^+$: 256.0928
found: 256.0926.

4.3. Synthesis of 5cadC



Scheme S7: Four step synthesis of 5cadC as performed in this work, adapted from the literature.[11,12]

4.3.1. Synthesis of silyl protected 5'-carboxyethyl-2'-deoxy-cytidine (**5**)

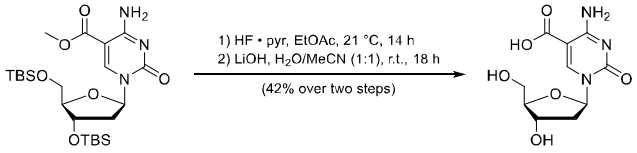


According to a modified standard literature procedure:[11]
In a high-pressure reactor, silyl protected 5-iodo-2'-deoxycytidine **3** (400 mg, 0.688 mmol, 1.00 equiv.) and bis(acetonitrile)palladium dich
dissolved in anhydrous methanol (8.5mL). The reactor was flushed with carbon monoxide gas (3.5 bar) while stirring the reaction solution
This flushing procedure was repeated two more times to achieve carbon monoxide saturation of the reaction solution. A solution of *N,N*-di-*i*
equiv.) in anhydrous methanol (0.25 mL) was added *via* syringe in one portion, the reactor was filled with carbon monoxide gas (3.5bar) ar
After 16 h, the reaction was cooled to 21 °C, the pressure was released and the orange suspension was filtered through a plug of silica. The
acetate (100 mL) and the filtrate was concentrated *in vacuo*. The resulting oily orange crude product was combined with the crude product of
purified by flash column chromatography (33% to 50% to 67% ethyl acetate in hexanes) to afford silyl-protected 5-methoxycarbonyldeoxy
yield) as a slightly yellow powder.

^{1}H NMR (400 MHz, $CDCl_3$):
δ = 8.58 (s, 1H), 7.88 (s, 1H), 6.12 (dd, *J* = 7.3, 5.8 Hz, 1H), 5.81 (s, 1H), 4.29 (dt, *J* = 6.3, 2.4 Hz, 1H), 4.00 (q, *J* = 2.9 Hz, 1H), 3.79 (dd, *J* =
= 11.3, 3.1 Hz, 1H), 2.56 (ddd, *J* = 13.5, 5.9, 2.5 Hz, 1H), 1.91 (ddd, *J* = 13.4, 7.4, 6.1 Hz, 1H), 0.82 (d, *J* = 6.7 Hz, 18H), 0.01 (s, 3H), 0.01 (

HR-MS (ESI):
calculated for $C_{23}H_{44}N_3O_6Si_2$ $[M+H]^+$: 514.2763
found: 514.2767.

4.3.2. Synthesis of 5-carboxy-2'-deoxycytidine (5cadC)



5 **5cadC**

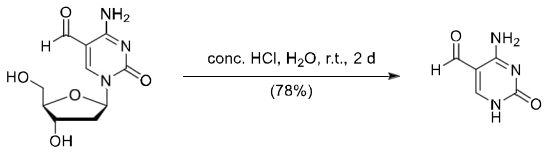
According to a modified standard literature procedure:[12]
Silyl-protected 5-methoxycarbonyldeoxycytidine **5** (520 mg, 1.01 mmol, 1.00 equiv.) was suspended in anhydrous ethyl acetate (30 mL) in a
pyridine complex (70 wt% HF, 0.40 mL, 15 mmol, 15 equiv.) was added whereupon a clear solution resulted. After 14 h, the reaction suspens
silane (4.5 mL, 32 mmol, 32 equiv.) was added slowly *via* syringe. After 3 h, the reaction was allowed to warm to room temperature and was
suspension was diluted with methanol and all volatiles were removed *in vacuo*. The yellow crude product was dissolved in a mixture of ace
hydroxide (290 mg, 12.1 mmol, 12.0 equiv.) was added. After 18 h, the suspension was acidified to pH 3 with 3.0 M hydrochloric acid and
856mg of a dark brown solid. A portion of 227 mg was purified by HPLC (Column: Agilent Prep-C18 column 50×10.0mm; 5 µm; 2% MeCN in
0.1% trifluoracetic acid) to afford 5-carboxy-2'-deoxycytidine (**5cadC**) (30.6 mg, 0.113 mmol, 42% extrapolated yield over two steps) as a co

^{1}H NMR (400 MHz, D_2O):
δ = 8.94 (s, 1H), 8.49 (s, 1H), 6.25 (t, *J* = 6.1 Hz, 1H), 4.48 (dt, *J* = 6.5, 4.6 Hz, 1H), 4.13 (td, *J* = 4.5, 3.3 Hz, 1H), 3.91 (dd, *J* = 12.6, 3.3 Hz,
(ddd, *J* = 14.2, 6.6, 4.9 Hz, 1H), 2.42 (dt, *J* = 14.2, 6.1 Hz, 1H) ppm.

HR-MS (ESI):
calculated for $C_{10}H_{14}N_3O_6$ $[M+H]^+$: 272.0877
found: 272.0875.

4.4. Synthesis of nucleobases 5-formylcytosine (5fC) and 5-carboxycytosine (5caC)

4.4.1. Synthesis of 5-formylcytosine (5fC)



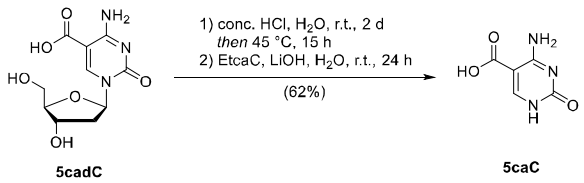
5fdC **5fC**

5-Formyl-2'-deoxycytidine (5fdC; 119 mg, 0.446 mmol, 1.00 equiv.) was suspended in water (7.0 mL) and concentrated hydrochloric acid (37
stirred at room temperature open to air for two days and then neutralized with 25% aqueous ammonia solution. The mixture was diluted with
The crude product was purified by HPLC (Column: Agilent Prep-C18 column 50×10.0mm; 5 µm; 100% H_2O for 4 min, MeCN) to afford 5-form
as a slightly tan solid.

^{1}H NMR (400 MHz, DMSO, 295 K):
δ = 9.49 (s, 1H), 8.50 (s, 1H), 8.49 (s, 1H).

HR-MS (ESI):
calculated for $C_5H_4N_3O_2$ $[M–H]^-$: 138.0309.
found: 138.0308.

4.4.2. Synthesis of 5caC



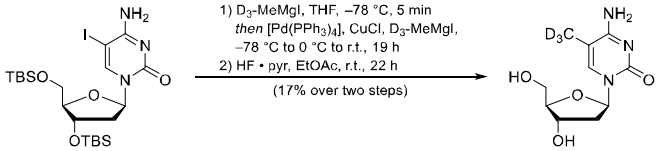
5cadC 5caC

5-Methoxycarbonyl-2'-deoxycytidine (5cadC; 49.2 mg, 0.172 mmol, 1.00 equiv.) was suspended in water (2.6 mL) and concentrated hydro
resulting clear solution was stirred at room temperature open to air for 2 days, then the mixture was heated to 45 °C. After 15 h, the reacti
diluted with methanol (5 mL) and concentrated *in vacuo* to afford a yellowish solid.
This crude product was combined with 5-ethoxycarbonyl-2'-deoxycytidine (37.3 mg, 0.375 mmol combined, 1.00 equiv.) and suspended i
4.50 mmol, 12.0 equiv.) was added upon which a clear solution resulted. After stirring at room temperature for 24 h, the solvent was remove
HPLC (Column: Agilent Prep-C18 column 50×10.0mm; 5 µm; 100% H_2O for 4 min, H_2O contained 0.1% trifluoracetic acid) to afford 5-carboxy

^{1}H NMR (400 MHz, DMSO, 295 K):
δ = 8.24 (s, 1H).

HR-MS (ESI):
Calculated for $C_5H_4N_3O_3$ $[M–H]^-$: 154.0258
Found: 154.0257

4.5. Synthesis of D_3-5mdC



3 D$_3$-5mdC

According to a modified literature procedure:[13]
5-Iodo-2'-deoxycytidine **3** (302 mg, 0.518 mmol, 1.00 equiv.) was dissolved in anhydrous tetrahydrofuran (11 mL) and the solution was
methylmagnesium iodide in diethylether (0.52 mL, 0.52 mmol, 1.0 equiv.) was added dropwise *via* syringe and the clouded solution stirred for
(59.9 mg, 51.8 □mol, 0.100 equiv.) and copper(I) chloride (256 mg, 2.59 mmol, 5.00 equiv.) were added, followed by more D_3-methylma
5.0 equiv.). The flask was covered with aluminum foil and the reaction mixture stirred at –78 °C for another 50 min. Then the reaction was a
during which it was allowed to warm to room temperature. After 18 h, the yellowish-green suspension was diluted with dichloromethane (30
solution (15 mL) and water (15 mL). The layers were separated and the aqueous layer extracted with dichloromethane (3×40 mL). The
saturated aqueous sodium chloride solution (50 mL) and dried over sodium sulfate. The dried solution was filtered and the filtrate was co
purified by flash column chromatography (0% to 9% to 19% methanol in dichloromethane + 1% triethylamine) to afford an inseparable mixture
protected D_3-5-methyl-2'-deoxycytidine as a yellowish solid.
This mixture was then diluted in anhydrous ethyl acetate (15 mL) in a plastic tube in open air and hydrogen fluoride pyridine complex (70% H
dropwise *via* syringe. The yellowish solution was stirred at room temperature overnight. After 22 h, the resulting suspension was cooled
16 mmol, 32 equiv.) was added. After 30 min, the reaction mixture was allowed to warm to room temperature and was stirred for another
methanol (10 mL) and concentrated *in vacuo*. The crude product was purified by HPLC (Column: Agilent Prep-C18 column 50×10.0mm; 5
H_2O both contained 0.1% trifluoracetic acid) to afford D_3-5-methyl-2'-deoxycytidine (D_3-5mdC; 21.5 mg, 88.0 µmol, 17%) as a colorless solid

^{1}H NMR (400 MHz, D_2O):
δ = 7.95 (s, 1H), 6.28 (t, *J* = 6.5 Hz, 1H), 4.48 (dt, *J* = 6.5, 4.2 Hz, 1H), 4.10 (dt, *J* = 4.9, 3.8 Hz, 1H), 3.89 (dd, *J* = 12.6, 3.4 Hz, 1H), 3.79 (dd

^{2}H NMR (61 MHz, H_2O):
δ = 1.98 (s, 3H).

^{13}C NMR (101 MHz, D_2O):
δ = 159.6, 148.8, 141.5, 103.7, 87.0, 86.2, 70.1, 60.9, 39.1.

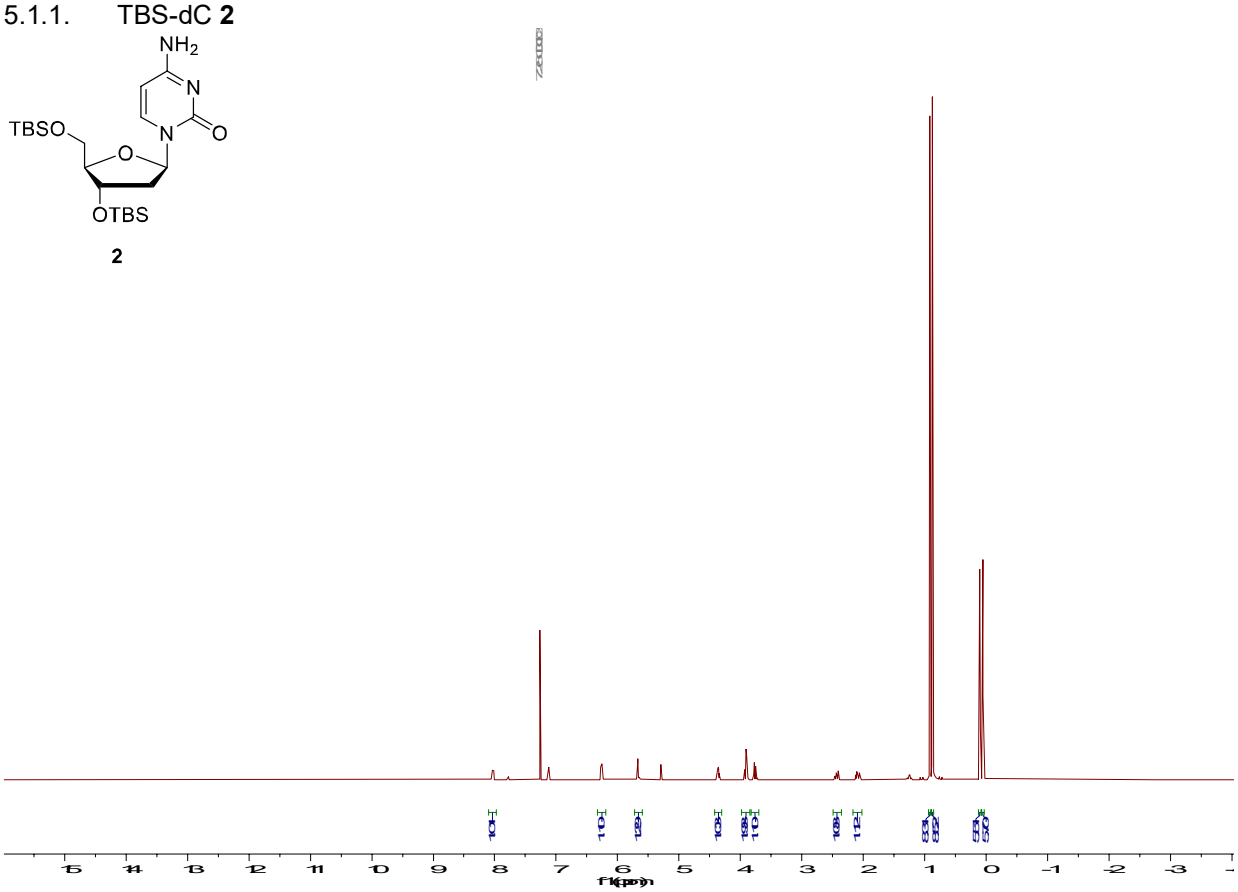
HR-MS (ESI):
Calculated for $C_{10}H_{12}D_3N_3NaO_4$ $[M+Na]^+$: 267.1143
Found: 267.1139

5. Appendix

5.1. NMR spectra

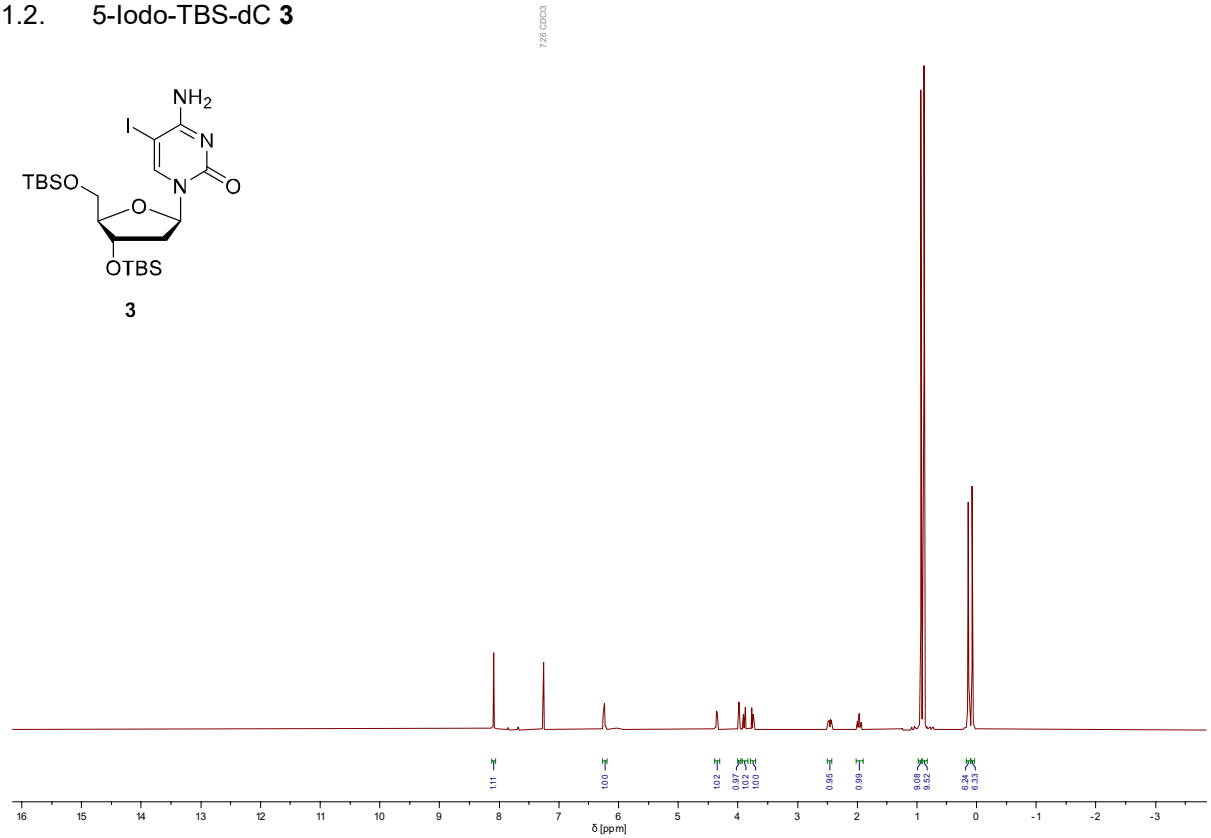
5.1.1. TBS-dC **2**

2



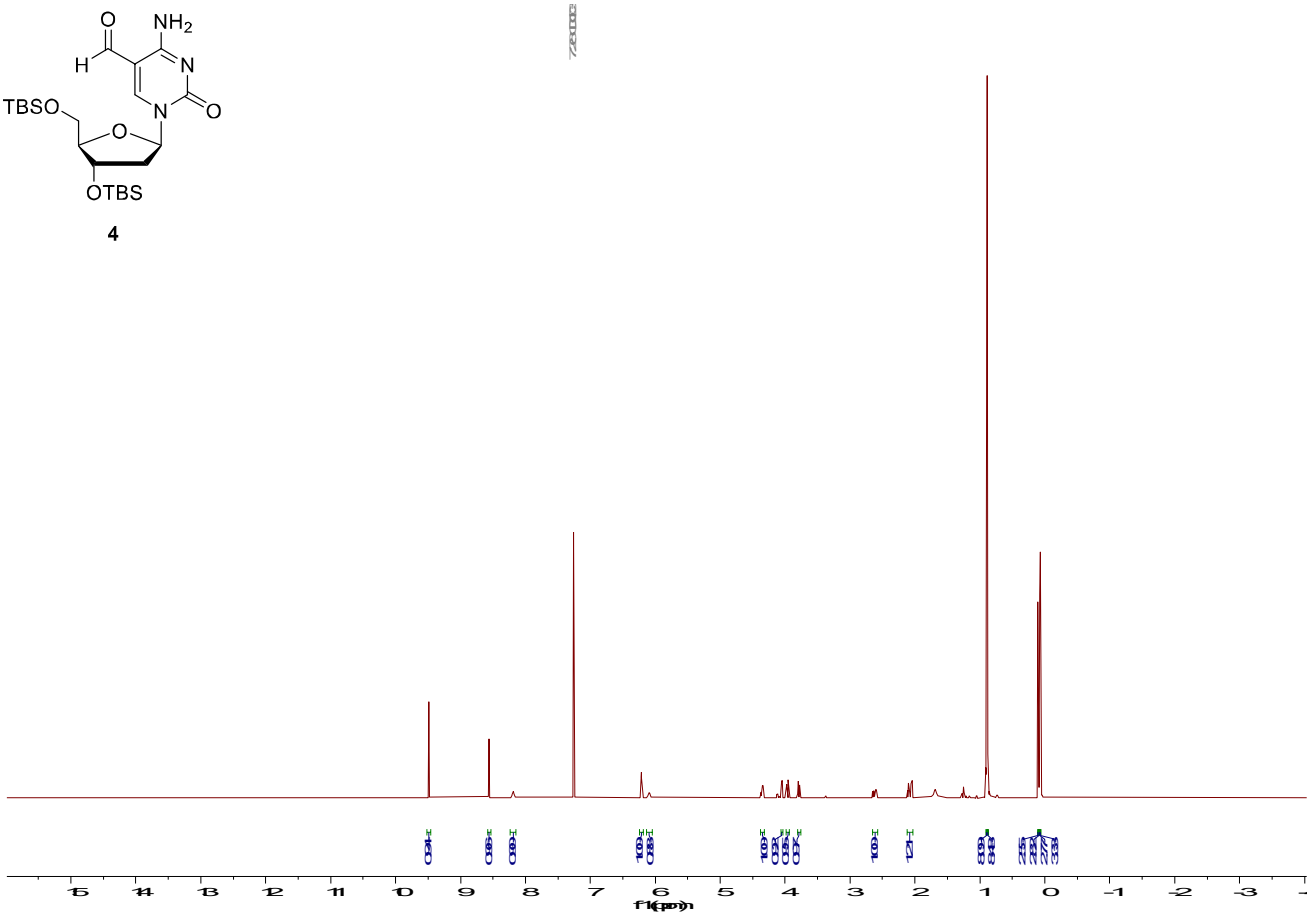
5.1.2. 5-Iodo-TBS-dC **3**

3

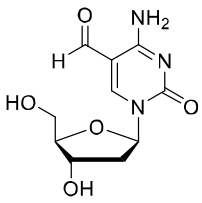


5.1.3. 5-Formyl-TBS-dC 4

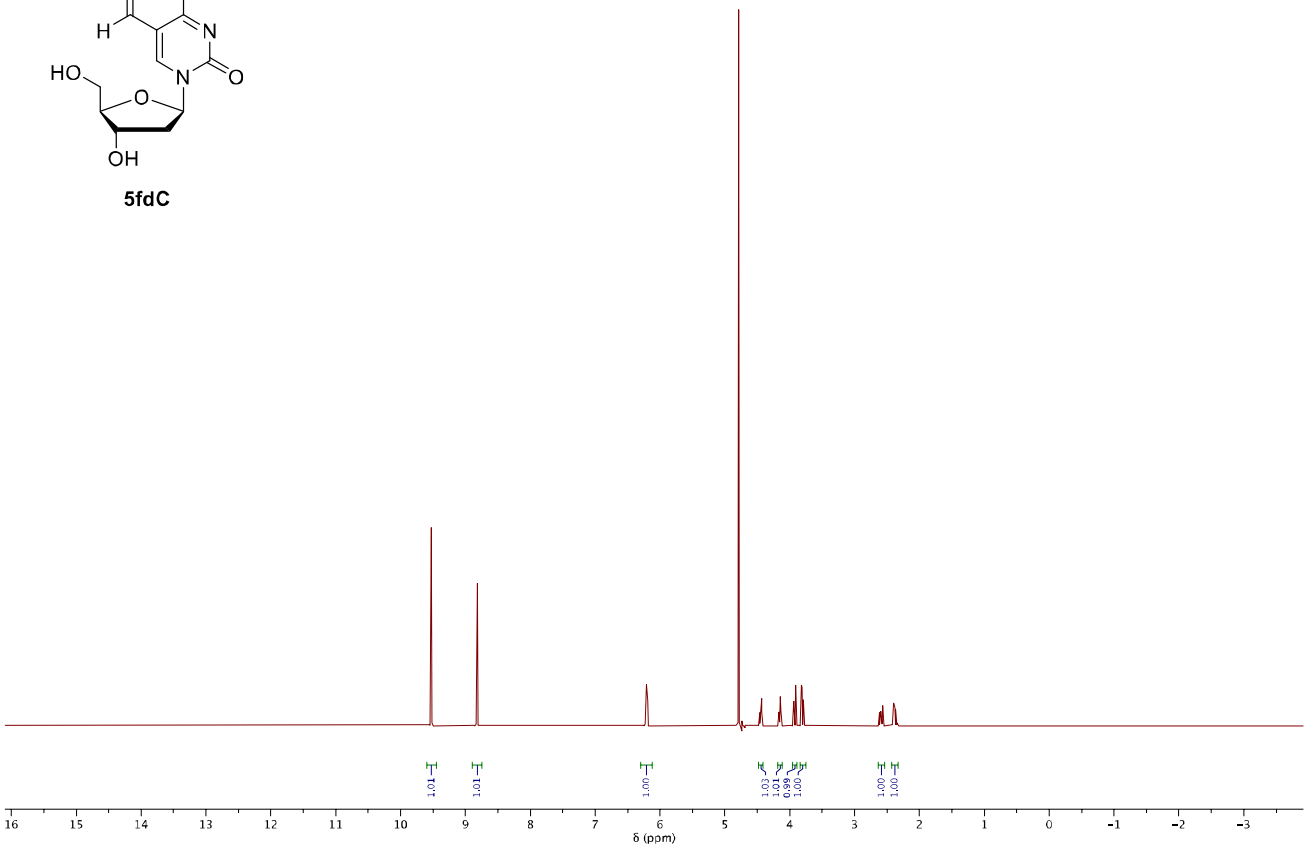
4



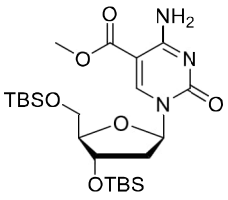
5.1.4. 5-Formyl-2'-deoxycytidine (5fdC)



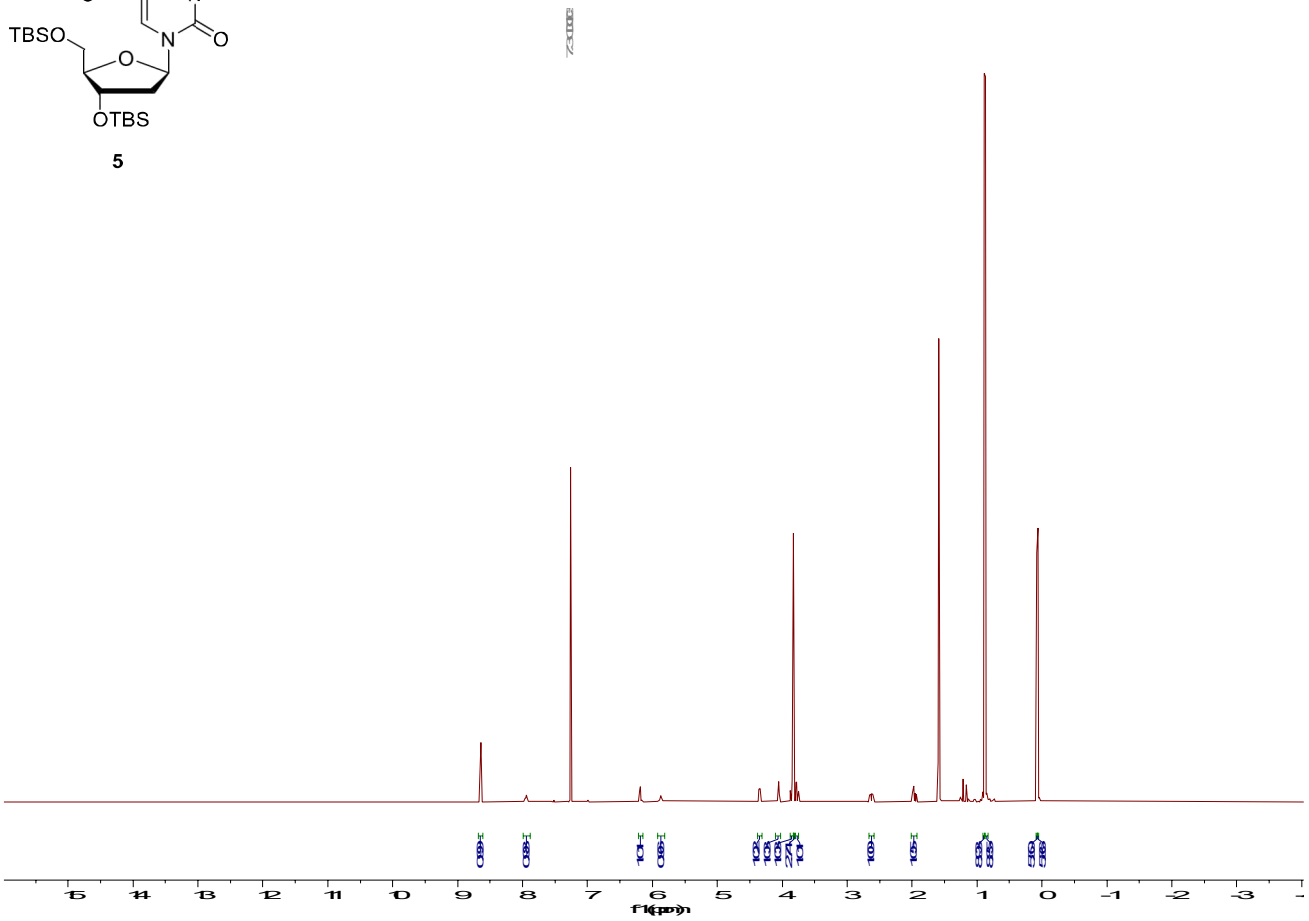
5fdC



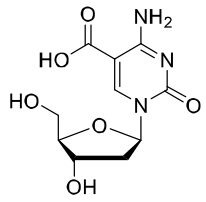
5.1.5. 5-(Methoxycarbonyl)-TBS-dC **5**



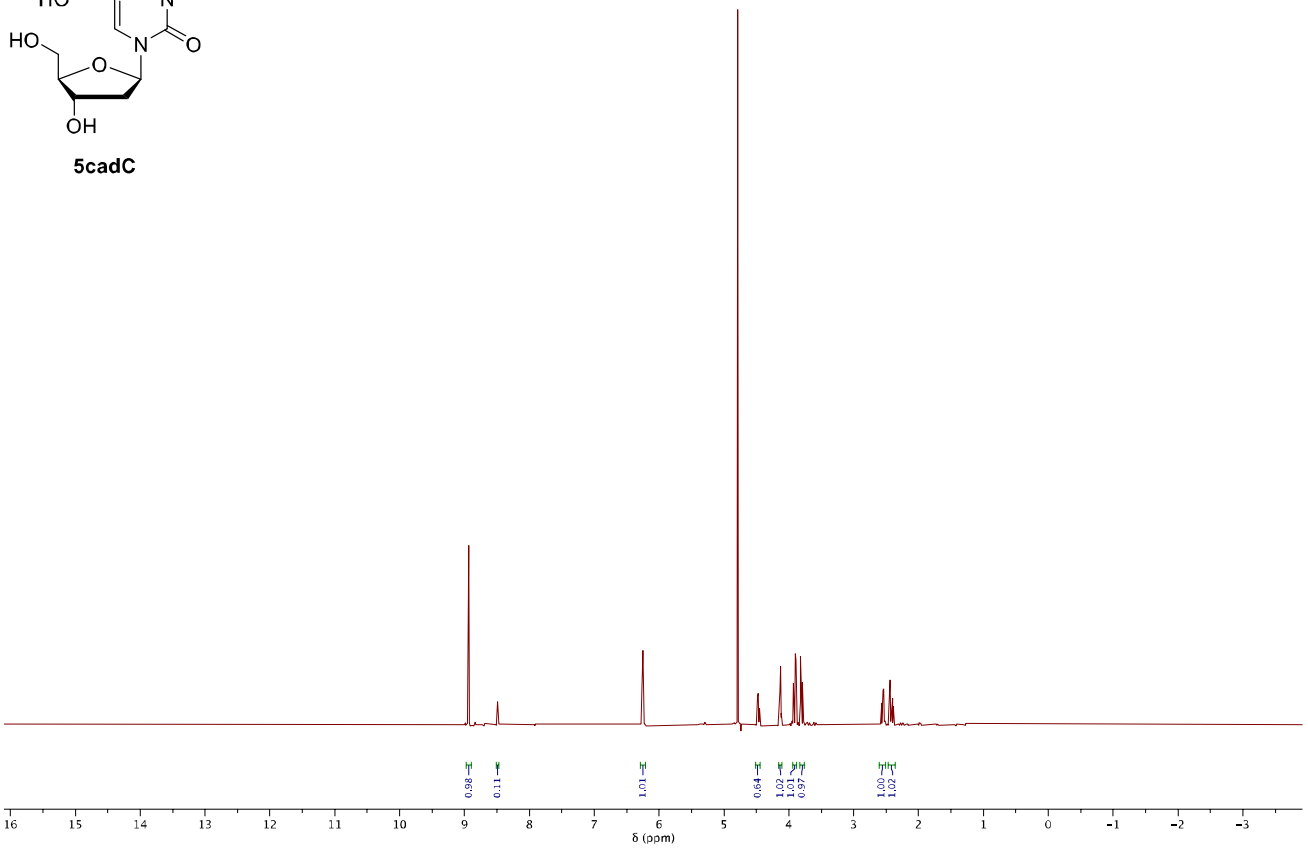
5



5.1.6. 5-carboxyl-2'-deoxycytidine (5cadC)

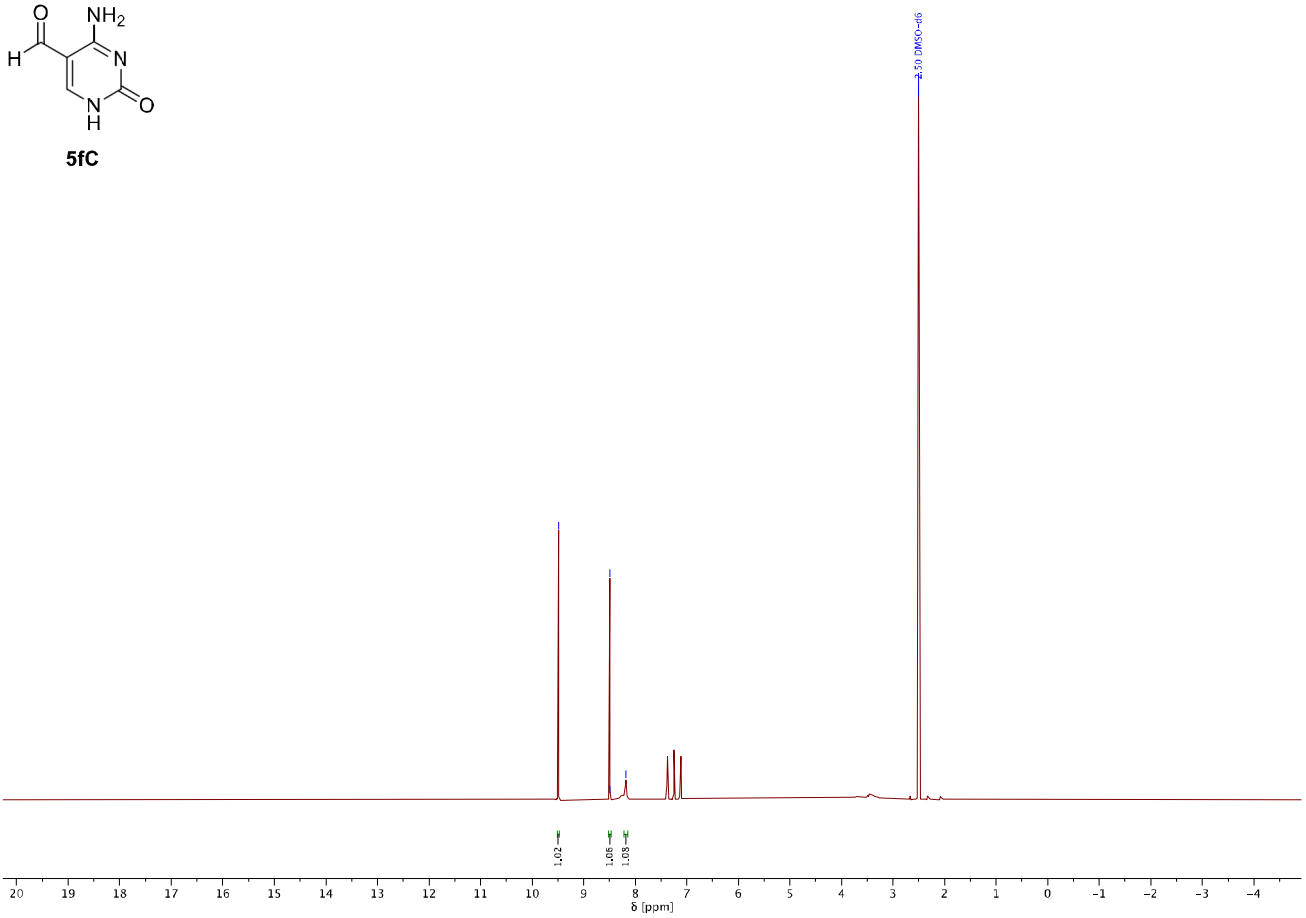


5cadC

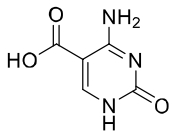


5.1.7. 5-Formylcytosine (5fC)

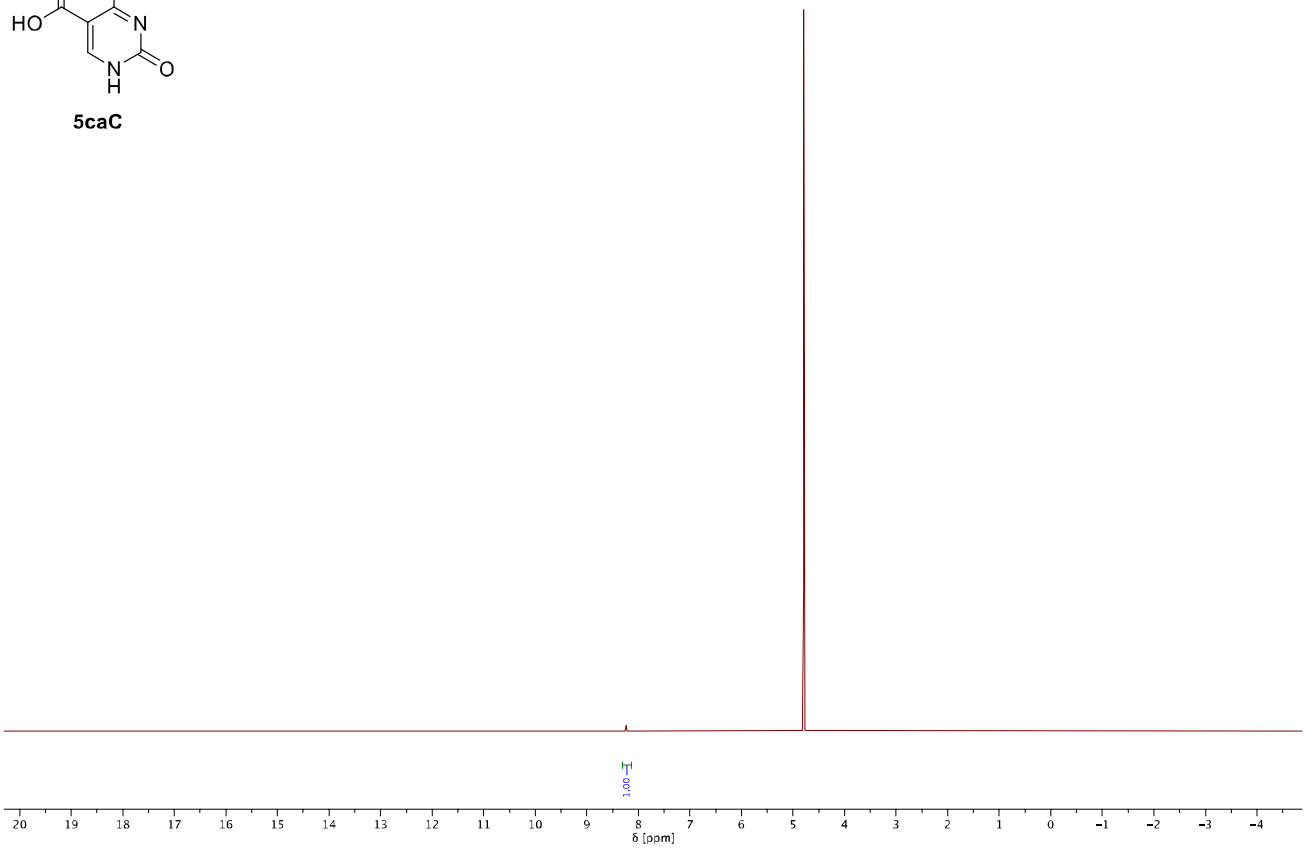
5fC



5.1.8. 5-carboxylcytosine (5caC)

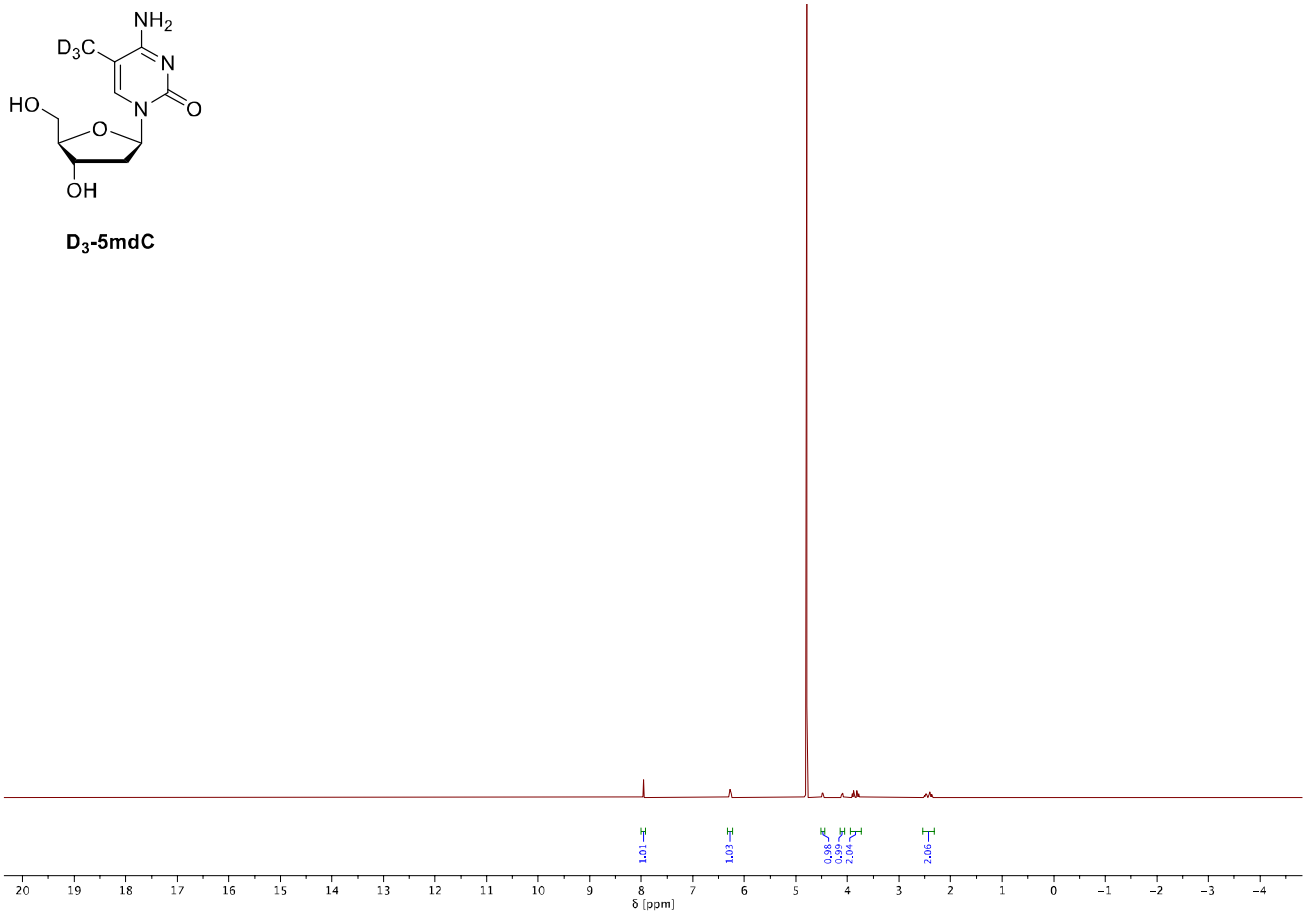


5caC



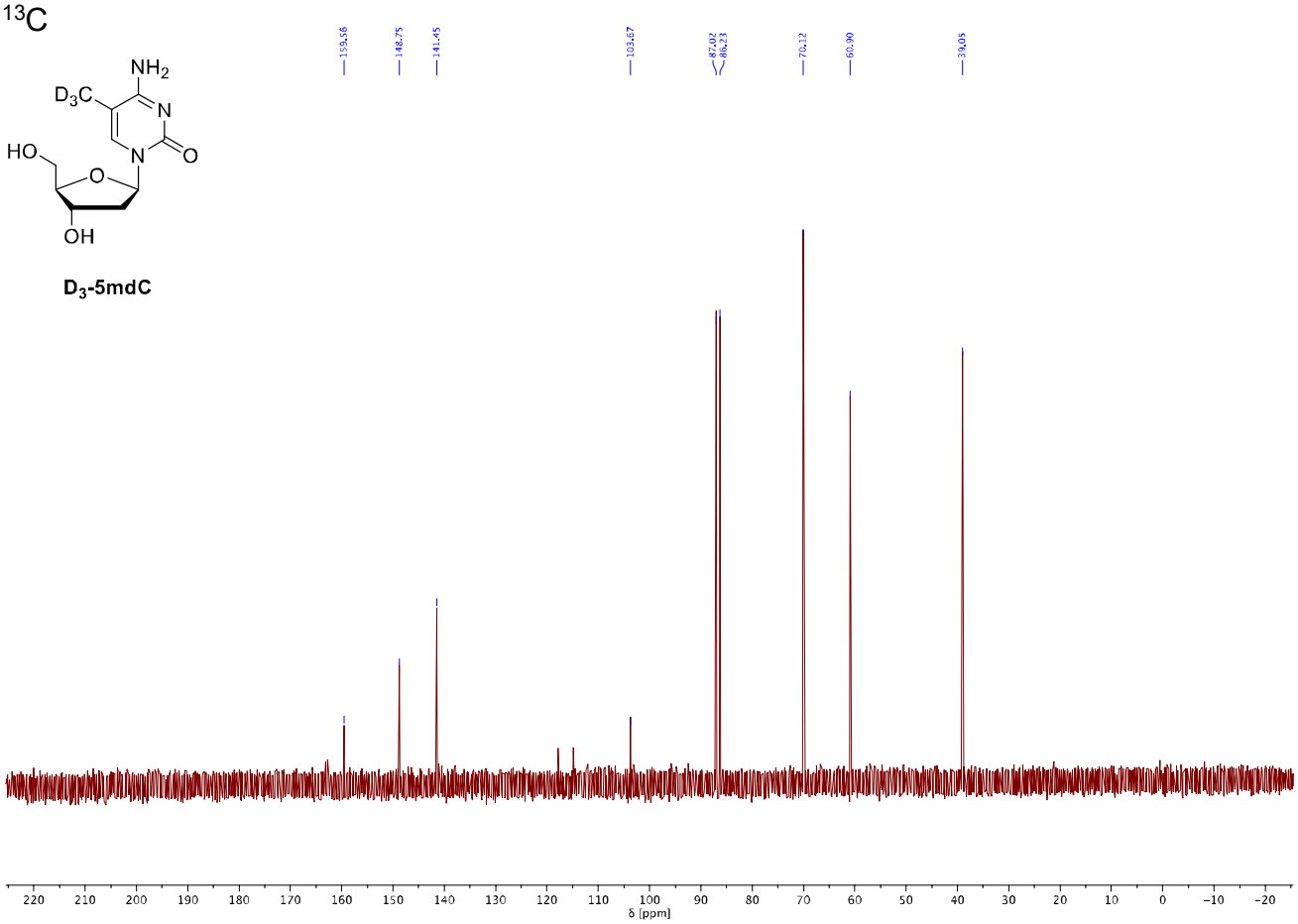
5.1.9. D_3-5-methyl-2'-deoxycytidine (D_3-5mdC)

^1H



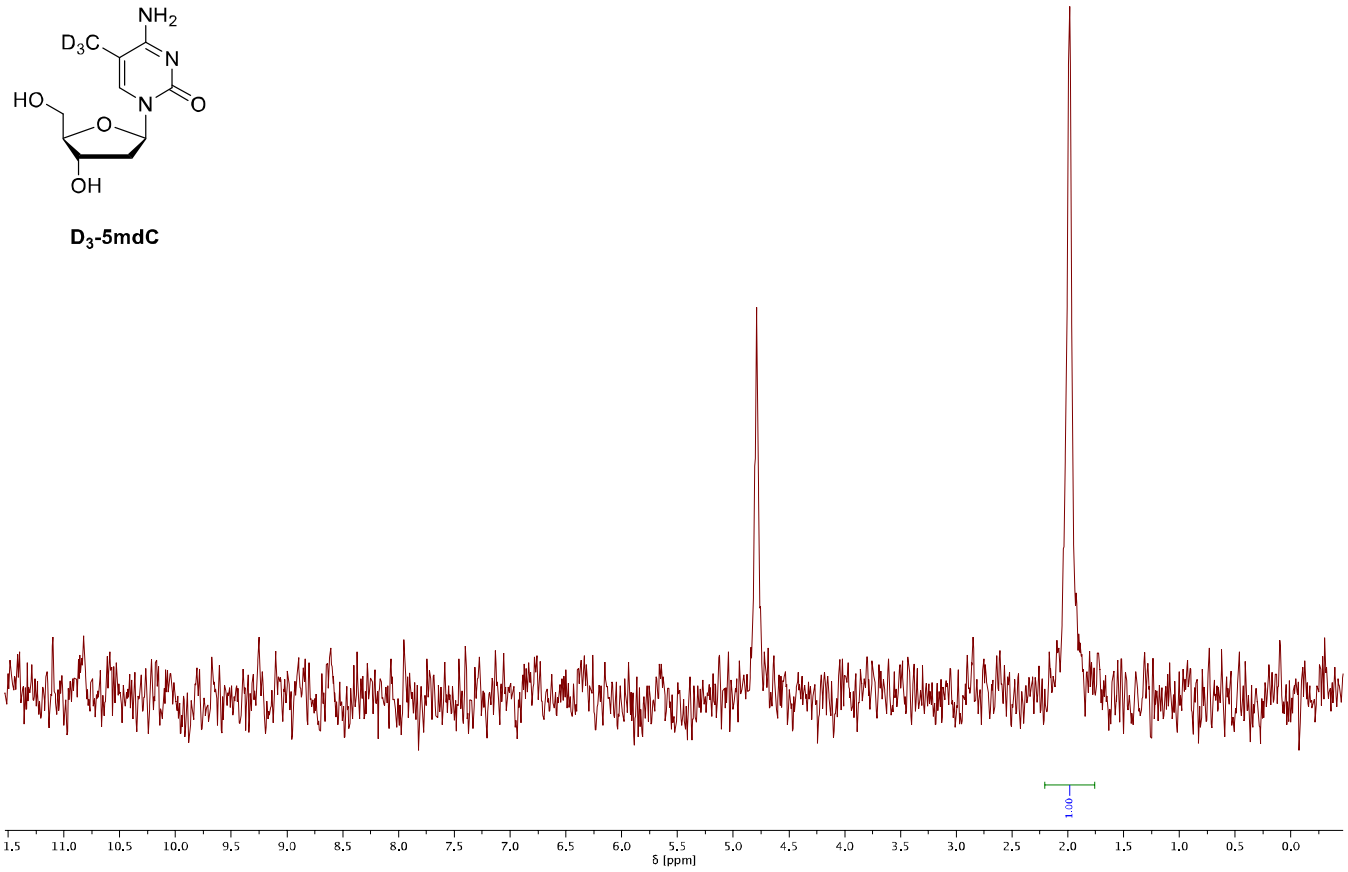
^{13}C

D_3-5mdC



^{2}H

D_3-5mdC



References

[1] G. R. Fulmer, A. J. M. Miller, N. H. Sherden, H. E. Gottlieb, A. Nudelman, B. M. Stoltz, J. E. Bercaw, K. I. Goldberg, *Organometallics* **2010**, *29*, 2176–2179.
[2] F. R. Traube, S. Schiffers, K. Iwan, S. Kellner, F. Spada, M. Müller, T. Carell, *Nat. Prot.* **2019**, *14*, 283–312.
[3] L. Hu, J. Lu, J. Cheng, Q. Rao, Z. Li, H. Hou, Z. Lou, L. Zhang, W. Li, W. Gong, M. Liu, C. Sun, X. Yin, J. Li, X. Tan, P. Wang, Y. Wang, D. Fang, Q. Cui, P. Yang, C. He, H. J
[4] S. Steenken, S. V. Jovanovic, L. P. Candeias, J. Reynisson, *Chem. Eur. J.* **2001**, *7*, 2829–2833.
[5] N. S. W. Jonasson, L. J. Daumann, *Chem. Eur. J.* **2019**, *25*, 12091–12097.
[6] T. Chantarojsiri, Y. Sun, J. R. Long, C. J. Chang, *Inorg. Chem.* **2015**, *54*, 5879–5887.
[7] J. Cong, D. Kinschel, Q. Daniel, M. Safdari, E. Gabrielsson, H. Chen, P. H. Svensson, L. Sun, L. Kloo, *J. Mater. Chem. A* **2016**, *4*, 14550–14554.
[8] D. B. Gophane, S. T. Sigurdsson, *Beilstein J. Org. Chem.* **2015**, *11*, 219–227.
[9] M. Münzel, D. Globisch, C. Trindler, T. Carell, *Org. Lett.* **2010**, *12*, 5671–5673.
[10] D. Globisch, M. Münzel, M. Müller, S. Michalakis, M. Wagner, S. Koch, T. Brückl, M. Biel, T. Carell, *PLoS One* **2010**, *5*, e15367.
[11] M. Münzel, U. Lischke, D. Stathis, T. Pfaffeneder, F. A. Gnerlich, C. A. Deiml, S. C. Koch, K. Karaghiosoff, T. Carell, *Chem. Eur. J.* **2011**, *17*, 13782–13788.
[12] B. Steigenberger, S. Schiesser, B. Hackner, C. Brandmayr, S. K. Laube, J. Steinbacher, T. Pfaffeneder, T. Carell, *Org. Lett.* **2013**, *15*, 366–369.
[13] M. Münzel, D. Globisch, T. Brückl, M. Wagner, V. Welzmiller, S. Michalakis, M. Müller, M. Biel, T. Carell, *Angew. Chem. Int. Ed.* **2010**, *49*, 5375–5377.